

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:34:58 ; Search time 14.65 Seconds
(without alignments)
1188.768 Million cell updates/sec

Title: US-09-905-088A-245

Perfect score: 3732

Sequence: 1 MRLVAPLLAMVAGATATV.....RKLPSSSEGITLPLLSQNS 713

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1. Issued_Patents_AA:*
2. /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
3. /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
4. /cgn2_6/prodata/2/1aa/5C.COMB.pep:*
5. /cgn2_6/prodata/2/1aa/5D.COMB.pep:*
6. /cgn2_6/prodata/2/1aa/5E.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1656.5	44.4	708	4	US-09-131-648-2
2	442	11.8	1091	3	US-08-986-485-5
3	408	10.9	1101	3	US-08-986-485-2
4	407.5	10.9	605	1	US-08-190-802A-49
5	407.5	10.9	605	4	US-08-473-089-49
6	407.5	10.9	605	4	US-08-473-089-49
7	402	10.8	673	4	US-09-063-950-2
8	397.5	9.5	1523	4	US-09-182-024A-2
9	355.5	9.2	603	1	US-08-190-802A-50
10	344	9.2	603	4	US-08-473-089-50
11	344	9.2	603	4	US-08-473-089-50
12	331	8.9	1525	3	US-09-191-647-2
13	331	8.9	1525	3	US-09-191-647-2
14	331	8.9	1525	3	US-09-191-647-2
15	331	8.9	1525	3	US-09-191-647-2
16	331	8.9	1525	3	US-09-191-647-2
17	331	8.9	1525	3	US-09-191-647-2
18	331	8.9	1525	3	US-09-191-647-2
19	331	8.9	1525	3	US-09-191-647-2
20	331	8.9	1525	3	US-09-191-647-2
21	331	8.9	1525	3	US-09-191-647-2
22	331	8.9	1525	3	US-09-191-647-2
23	331	8.9	1525	3	US-09-191-647-2
24	331	8.9	1525	3	US-09-191-647-2
25	331	8.9	1525	3	US-09-191-647-2
26	331	8.9	1525	3	US-09-191-647-2
27	331	8.9	1525	3	US-09-191-647-2

28	273	7.3	1112	4	US-09-353-585-3	Sequence 3, Appl1
29	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
30	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
31	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
32	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
33	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
34	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
35	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
36	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
37	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
38	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
39	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
40	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
41	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
42	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
43	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
44	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1
45	269	7.2	353	4	US-09-353-585-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-131-648-2
Sequence 2, Application US/09131648
Patent No. 6168920
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Giegler, Karl J.
TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
FILE REFERENCE: PE-0576 US
CURRENT APPLICATION NUMBER: US/09/131,648
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PERL Program
SEQ ID NO 2
LENGTH: 708
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: 2687731
US-09-131-648-2

Query Match 44.4%; Score 1656.5; DB 4; Length 708;
Best Local Similarity 47.2%; Pred. No. 2.6e-144;
Matches 333; Conservative 109; Mismatches 217; Indels 47; Gaps 5;

QY	17	TATVPVPMHVCPPQACQIRFWYTPRSSYREATVDCNDIFLAVPALPAGQITLL	76
DB	17	TTLVQAVDKKVCPCRLCTCEIRFWFTPRISYMEASTVDCNDIGLTFPAPLPAQITLL	76
QY	77	QNSIYRVQSEGLYANLTELDELQNSFSPARDCEFHLPOLLSLHLENOITRLSDHS	136
DB	77	QTNMIAKIEYS-TFPYVNLGLDLSQNMISVTVINVKMPQLSYLLENNKITLPEKPC	135
QY	137	FAGIASLOELIYNHOLYRIAPRAFGSLNLRHLNSNLRALRDISWEPDMLEITLMI	196
DB	136	LSEISNLOELIYNHOLYRIAPRAFGSLNLRHLNSNLRALRDISWEPDMLEITLMI	195
QY	197	GGNVVAIIDNNFRPLANLRSILVLAGMMLREISDYALEGLOSLSISFYNDQIANVPRRA	256
DB	196	GEMTIRIKDMFKFLINLSLVINGINLITFIPNALVLENTLSISFYNDQIANVPRRA	255
QY	257	LEQVPGKFLDLNKNVLOVRGPGDFANMLHKLKELGNNKEELVSTDFALVNPPELLTLD	316
DB	256	LQKVNKFLDLNKNVLRIRGDSNNMLHKLKELGNNKEELVSTDFALVNPPELLTLD	315
QY	317	ITNPRLSFIRHRAVHLPMQETLMNNALSLALHOQTVESIPMLQEVJLHGNPITKDCV	376

Db	316	ATNNRLSYIHNAFFELKRLKSLMLNSMLSLALYIGTIESSLPMLKEISHSPICRCOV	375
Qy	377	IMANATGRVRFLEPOSTLCAEPDLDLORPLVEEVPFPEKMDHCLPLISRFPSILOVA	436
Db	376	IRKMNKNTNTRFEEPPSLFCVDPPEEGGVNBOVHFRRMDEICLPLIAESPENLAVE	435
Qy	437	SGESVYLHCRAALAEPEELIYVTPAGRLPLPAAGHGRVYVPEGLTLRFVYAEAGLYT	496
Db	436	AGSYVSEFCRCATAPDPELIYTWTPESGRLPLNTLTKFVYHSEGTLDINCVYPKGGIYT	495
Qy	497	CYANQVIGADPKTVSVVVGRRALLQGRDEGQGLKLVQDETPIYHLLSWTPENYVNTL	556
Db	496	CIATRLVADLKSVMKIKGVGSPD---DNNGSLNIRIDQANSVLVSMWASSKILIKSSV	552
Qy	557	YWSSASLIRGGCATALARLPKGTSHSYNTRILLQATEWACLOVAFAADHOTALACWATK	616
Db	553	KTATAVYKTENSHAAQSARLPSPVYKYNLTPLNLTSTYEKLCIDIFTIYKKNRKKCVNTYK	612
Qy	617	-----EATSCRRALCDNRGLATLAILALAVLLAAGIAAHLTGQPRKGVGRR-	663
Db	613	GLHPDQEKYEKNNTITMLACGLGLGIIIGVICISCL-----SPENKCDGHS	660
Qy	664	-----PLPAPAAEWGMSAPSVRYGAPVILPMN	691
Db	661	YVRNTLQKPTPLAGELIYPLPLIMLWDKKEKSTSLKAKKATVIGLPTN	706

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	11.8%;	Score 442;	DB 3;	length 1091;
Best Local Similarity	23.2%;	Pred. No. 1.1e-31;		
Matches 169;	Conservative 106;	Mismatches 255;	Indels 198;	Gaps 22;

[illegible]

```

STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,485
FILING DATE: 08-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/059,448
FILING DATE: 22-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1101 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-986-485-2

Query Match      10.9%: Score 408; DB 3; Length 1101;
Best local similarity 21.8%; Pred No. 1,5e-28;
Matches 185; Conservative 124; Mismatches 277; Indels 264; Gaps 30;

QY 8 LLLAVY-----AATITVYVYVPMVPPPOCACQIRPMTTSSRYRATTVQCN 56
DB 18 LLLMLLLRLREPYTAASPR-----PCMAACTAGDPTCAGDS-----LDCG 62
QY 57 DLFVAVPALPACTOTLLQSNSTVRYDOSLGITLANLT-----LDCG 97
DB 63 GGLALALPGDLPFTSLSLSLNKLAEIDRAGFEDLPNLOEYVNNHETLANVASLGAGSS 122
QY 98 -----LDLSONSFSDARCDPRHALPOLLSLHLENOQL 129
DB 123 QVVALLFQQQNNSLDGSOLKAYLSLEVLLNLTNNITTEVRNTYPPHGPPIKELNLGAKRI 182
QY 130 TLELHSPFAGLA-SLOELYLNHNOYRIAPRAFSGLSNLRLHNSLRLAIDSRWFEML 188
DB 183 GTLELGAFDGSLSLTLTLRLSKNNRTOLPVPRAK-LPRLTQDLNRRRIHLEGLTFEQL 241
QY 189 PULELLMIGKNKYDAILDNFRPLANLRSVLAGMUREISDYALEGLQSLSELSFYDQ 248
DB 242 NSLEVLKLDORNNISKLTLDGAFWGLSKMIVLHLEYDSLVEVNSGYLLTLHLQHLNNS 301
QY 249 LARVRR-----ALEQVPGKFLDNKNPDLORVPGDPPANM 284
DB 302 IARIHRKQWSECKLHELVLSPNNLTRLDEESLAEISLSYLRSHNSISHLEAGFGL 361
QY 285 LALKELGLNNEELVSID----KFAVNLPELTKDITNNRSLFIRAFHHLDPOMETL 340
DB 362 RSLRVLDLDHNEISGTIEDTSGAFSGLEFGR-SKLTLEFGR-KIKSVAKRAFSGLEHHL 419
QY 341 MLNNNALSLALHQOTVESLPLNLOEVLGNGPIRCDCVYIR-----ANAT----- 383
DB 420 NIGGNARISVQFOAFKMKMLKEHLHSSDSFLCDQKMLPWLIGMNLGAFVTAACHP 479
QY 384 -----GTRVRFIEPOSTLC-----AEP----- 400
DB 480 ESJLKGQSLFVSPEPSFVCCDFLKPQIITQPEITMANVYGNDRIFTCASASSSSPMTFPAK 539

```

```

QY 401 -----PDIGR-----LFEVREVP-----R 414
DB 540 KDNEVLTNADMENEFVNHADGEVMEYTTLLRLQVTEGHEGRYOCVITNHFSGTYSNKA 599
QY 415 EMDHCLPLISPSRSPPSLOVAGSESMVLACRALAEPEPIYVWPYAGIRLTPAHAGRRY 474
DB 600 RLTVNVLVPS-STKTPRHDITIRTTVARLECAATGHNRPQIDAMOKXGDTP-PAARERRR 656
QY 475 RVPE-EGTELRVYAEAGLYTCVAQNLVG--ADTKVSVVYGRALLDOPGRD-----GQ 527
DB 657 HVMDDDVFFITVDKIDDAQVYSCIAQNSAGSISAMTLVLETPSLVYPLIEDRVVSVEE 716
QY 528 GLELRVQETHPRPHILLSWTPPTVSTNLTWSSASSLRGQAGATALARPRTGTHSYNTT-- 585
DB 717 TVALOCKATG-----NPP-----PRITW-----FKGDRPLST-----TERNHITPD 752
QY 586 -RL-----QATEYWACIQ--VAFADAPHTOLACVWARTKEATSCNRALGDRPGLIAI 634
DB 753 NQLLVQNVVADAGARYCEMSNTLGTERRAHLSQLSVL-----PAACG-RKDGTVVGIETI 806
QY 635 LALAVLLIAA 644
DB 807 AVVSIVLTS 816

RESULT 4
US-08-190-802A-49
Sequence 49, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fadian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0960
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor binding
INDIVIDUAL ISOLATE: protein complex, Fig. 32
US-08-190-802A-49

```

Query Match 10.9%; Score 407.5; DB 1; Length 605;

RESULT 6
US-08-473-089-49
Sequence 49, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Theeef
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor binding
INDIVIDUAL ISOLATE: protein complex, Fig. 32
US-08-473-089-49

Query Match 10.9%; Score 407.5; DB 4; Length 605;
Best Local Similarity 25.2%; Pred. No. 6,1e-29;
Matches 152; Conservative 72; Mismatches 173; Indels 205; Gaps 16;
QY 3 LVAPELLAMVA-----GATATVPRVPMHVPCCPQCACQIRPWTTPRSSY---REATT 52
DB 8 LALALLLSVALGPRSLLEGADPGTGEABGRCAPACVC-----SYDDADELS 57
QY 53 VDCNDLFTAVPALPAGTQTLLQNSIVRV-----DOSELTG----- 91
DB 58 VFCSRRMLTLPDQVPEGITALLMDGNNLSVPPAPQONISSIGFTLNLOGGQGLSLEPQA 117
QY 92 ---LANTELDSQNSFSFARDQDFHALPOLSLHSEENQTRLE----- 133
DB 118 LIGLENCCHLERNQRLSLATGTFATPALASIGSNNSLRLEDGLFEGLSLMDLN 177
QY 134 -----DHSFAGLASLOELVNLNQLYRIAPRAFSGSLNLRHLNLSNLLRAIDSR 184
DB 178 GWNLSAVLPAPAFRGSLRELVLAGNRLAYLDPALEFSGLAELRELDLSNALLRAIKANY 237
QY 185 FEMLPNLEIIMIGNKVDAT-----LDNN----- 208
DB 238 FVQJPRLOKXLDNRLAAAPGAFGLKALRWLDLSHNNAVAGLDETFPGGLGLRYLR 297
QY 209 -----FRPLANLSVLVAGNNLRISDVALEGLQSLSESISFTYNOLARY----- 252

DB 298 SHNAIASLRPRTFDHLFELEQGNRIKIRQLAERSPEGLGLEVLTLHNOLOEVKAGA 357
QY 253 -----PRALAEVPGKTLFLDNKNPLOWGVGDFFANMLHKELG 292
DB 358 FLGLTNAVMNLGNCRLNLPDQVFRGLKLSLHLEGSLGRIRPTFTGLSLRRL 417
QY 293 NNMEELVSDKFALVNLPELTKLDITNNPRLSFTHPRAFHLLPQMETIMNNALSAL-- 350
DB 418 KD-NGLVIERQSLWGLAELELDITSN-QTLPHRLFGGLKLEVLSSRRRLAE 475
QY 351 -----HQQTVESLPN-----LQEVGLHG 368
DB 476 DALGPIQRAFWLDVSHNRLEALPNSLLAPGLRLYLSLRNNSLRFTTPOPGLERLMEG 535
QY 369 NPICRDCVIR-----MANATGRV-RFTEP-----QSTLCAPPDQLRPLVR 409
DB 536 NPWDGCGPLKALRDFALQNSAVPRFVQAICEGDDCOPPAYTNNITCASPEVYGLDLR 555
QY 410 EV 411
DB 596 DL 597
RESULT 7
US-09-063-950-2
Sequence 2, Application US/09063950C
Patent No. 6225085
GENERAL INFORMATION:
APPLICANT: Holzman, Douglas A.
TITLE OF INVENTION: NOVEL LRSQ PROTEIN AND NUCLEIC ACID MOLECULES AND USES
FILE REFERENCE: MEI-019
CURRENT APPLICATION NUMBER: US/09/063,950C
CURRENT FILING DATE: 1998-04-21
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 673
TYPE: PRT
ORGANISM: Homo sapiens
US-09-063-950-2
Query Match 10.8%; Score 402; DB 4; Length 673;
Best Local Similarity 24.9%; Pred. No. 2,3e-28;
Matches 194; Conservative 73; Mismatches 306; Indels 206; Gaps 24;
QY 4 LVAPILLAMVAGATATVPRVPMHVPCCPQCACQIRPWTTPRSSYREATTVDQNDLFLAV 63
DB 7 LLLPILLALLAG-----PQVQGCPSGCOS-----QPQTVCTAROGTIV 46
QY 64 PPAALPAGTQTLLQNSIVRVDOSELTGYLANLTDELDSQNSFSARDQDFHALPOLSLH 123
DB 47 PPDVPTDVGTLVFENGITMDAGSFAGLPGLQLDLSQNDIASLPSVFPPLANLSLTD 106
QY 124 ILENDITLEDHSPAGLASLOELVNLNQLYRIAPRAFSGSLNLRHLNLSNLLRAIDSR 183
DB 107 LFLANLHETNTFTFGLRLERLYGKNRIRHIGGARDTIDRLLEKLONEDEALPPL 166
QY 184 WPEMLPNLEIIMIGNKVDAILDMMFRPLANLSVLVAGNNLRISDVALEGLQSLSELS 243
DB 167 ---RLPRLILLDLSHNSILA--LEPGILDPTANVEALRAGLGIGQODEGLFSRLRLHDLD 222
QY 244 FYDNOCLARYRRALQVYGLKFLDNKNPLOWGVGDFFANMLHKEGLNNMEELVSDK 303
DB 223 VSDNOLERP-----PVIR----- 236
QY 304 FALVNLPELTKLDITNNPRLSFTHPRAFHLLPQMETIMNNALSALHQQTVESLPNLOE 363
DB 237 ---GLRGTLRLRLAGNRIQALRPEDLAGLAALOELVSNLSIQALPGLDLSGLFPRLRL 292
QY 364 VGLHGMPICDCVIRMAN--ATGTRVRFIEPQSTLCAPPDQLRPLVREVPFRREM----- 416

```

Db 293 LAARPFNCVCPILSMFGWVRESHYTLASPEETRCFFPKNAGRLLLELDYADFQCPAT 352
QY 417 -TDHCLPLISPRSPPLSIQVASESMVILHORAIAEPEPELYWTWPAGLLTP----- 467
Db 353 TTTATVPTTTPVYREPTALSSSLAPTWLSFTAPATAPSPPTAPVTPVPOPODCPS 412
QY 468 -----AHAGRRYR---VYPEGTLRLRYTAEAGLYTCVQONLVGADTKTVSVVGRA 517
Db 413 TCLNGCTCHGTGRHNLACLPCEGF-----TGLY-CEGQ----- 444
QY 518 LLDPGDEGGLELRVOEHPHIIILSWTPPPTVSTNLTWSSASLRG-----QGAT- 570
Db 445 -----MGQTRSPPTVTP-----RPRSLTLGLEPVSPTSLRVGLORYLQSSSV 489
QY 571 -----ALALPRGTHSYNTIRLLQATEYWACL-----QYA 600
Db 490 QLNSLRLTYRNLSPDKRLVTLRLPASLAETVTLQRPNATVSCVMPLGPRVDEGEA 549
QY 601 FADAHQULACWARTKEATSCHRALGDRPGLAIIALAVLLAAGLAHLGTGPRKGVG 660
Db 550 CGEATTPPA---VSHNHAEVTOAREGNLPLLIA-PALAAVYLLAA-LAAVGAAYCYVRG-- 602
QY 661 GRRLPPPAWFMGWSAPSVRVVSAPLYL-----PNNPGRK-----LPRSSGETLL 706
Db 603 ---RAAAAAADKQGVPG---AGPLELEGVAVPLEPGKATGEGGALPSGSECEVPL 655

RESULT 8
US-09-063-950-5
; Sequence 5, Application us/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
; APPLICANT: Holzman, Douglas A.
; TITLE OF INVENTION: NOVEL LRGG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-019
; CURRENT APPLICATION NUMBER: US/09/063,950C
; CURRENT FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Papio hamadryas
US-09-063-950-5

```

```

Query Match 10.7%; Score 397.5; DB 4; Length 605;
Best Local Similarity 24.9%; Pred. No. 5,1e-28;
Matches 150; Conservative 73; Mismatches 173; Indels 207; Gaps 16;

QY 3 LLAVALLLAANA-----GATATVPPVPMHVPCCPQACQIRPWYTPRSSYREAT--T 52
Db 8 LALALLLSWVALGPRSLGAEPTPGAEAGPACATCAC-----SYDDEVNELS 57
QY 53 VVCNDLFLTAVPALPAGTQTLLOSNTIVR-----DQS 87
Db 58 VFCSSRNLTLPDGIPTGGQALWLDSSNNLSSIPPAFRNLSSLAFLNLQGGQLSLEPQA 117
QY 88 ELGYLANLTLELDLSONSFSDARDCDFHALPOLLSLHEENQTLRE----- 133
Db 118 LLG-LENICHLERNQRLSLAVGFATTPALALLGLSNNLSRLDEGLFGLGNLMDLN 176
QY 134 -----DHSFAGLASLOELVYNNOLYRIAPRAFSGLSNLRLHLNSNLRAIDSR 183
Db 177 LGWNSLAVLPAAAFRGISGLRELVLAGNRLAYLQPALESGIAELREIDLSSNALRAIKAN 236
QY 184 WPEMLPNTLEIIGGNKVDAI-----LDNM----- 208
Db 237 VFAQLPRLQKLYLDRNLIAAAYAPGAFGLKALRWLDLSHNVRAGILBDETPFGLLGLRVL 296
QY 209 -----FRPLANLRSIVLAGMNLREISDYALEGLSLESISFYDNLARY--- 252

```

```

Db 297 LSHNIAASLRRTBEDLHLELDLGNHRITQLARESPFEGIGQLEVYTLIDHNOLOEYKV 356
QY 293 -----PRALQVPGKLFKLDINKPQVGVGDEPANNMLHLEKG 291
Db 357 AFLGLTVAAVNNLGSNCRLNLPQGVFRGLKLSLHLEGSCIGRIRPHTFAGLSGLRRL 416
QY 292 LNNHEIVSIDKFLVYNLPETLTKDITNNPRLSTIHRARHHLQOMTLNNALSAL- 350
Db 417 LKD-NGLVIGESQSLWGLAELELDLTSN-QLTHLPOLFOGLGKLEYLLLSHRLAELP 474
QY 351 -----HOOTVESL-----PNLOEYGLH 367
Db 475 ADALGPLQRAFWLVDVSHNRLEALGSLASIGRLRYLNLNNSLRTPTPQPGELRLMLE 534
QY 368 GNPRLRCDCVIR---WANATGRV-RLEP-----QSTLCAEPDQLRLPV 408
Db 535 GNPWDCSCPLKALRDFALQNPASAVRFVQAICEGDDCQPPVYTYNNITCASPEVAGLDL 594
QY 409 REV 411
Db 595 RDL 597

RESULT 9
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
; APPLICANT: Connolly, Timothy
; APPLICANT: Rajput, Bhanu
; TITLE OF INVENTION: Human Silt Polypeptide and Polynucleotides Encoding
; FILE REFERENCE: 640100-271
; CURRENT APPLICATION NUMBER: US/09/182,024A
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/063,946
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/096,420
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-182-024A-2

```

```

Query Match 9.5%; Score 355.5; DB 4; Length 1523;
Best Local Similarity 21.8%; Pred. No. 1,8e-23;
Matches 154; Conservative 94; Mismatches 259; Indels 199; Gaps 18;

QY 13 VAGTATVPPVPMHVPCCPQACQIRPWYTPRSSYREATTYVCCNDLFLTAVPALPAGIQ 72
Db 18 LALALSVLSGPPRAVACPCTKC-----SAASVDCHGGLAVPGRGIPRNAE 64
QY 73 TLLQNSIYRVDOSEIGYLANLTLELDLSONSFSDARDCDFHALPOLLSLHEENQTLRL 132
Db 65 RLDDLNRNITRITRKMPAGIKNRLVHLIEDNOYSVIERAGFODLQLETLRLINKKLYL 124
QY 133 EDHSFAGLASLOELVYNNOLYRIAPRAFSGLSNLRLHLNSNLRAIDSRNPEMLPNE 192
Db 125 PELIFQSTPLTRLDLSENOIGIPKARFGITDVYNIQLDNNHISCIEDGAFALRLDLE 184
QY 193 ILMIGNKVDAIILDMNFRPLANLRSI----- 218
Db 185 ILLTNNNNISRIILVYSFNHMPKIRTLRLSHNHLVCDCHLAWLSDMLKORRVYGOFTLCMA 244
QY 219 --VLAGMNLREIS-----DYALEGL----- 236
Db 245 PVHLRGFNADVQKREYVCPAPRHSPPSCNANSISCPSPCTCSNNIIVDCRGKGLAEIFAN 304

```



```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MORASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 603 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor bind
INDIVIDUAL ISOLATE: pro. complex rat, Flg. 33
US-08-477-346-50

```

Query Match	9.28;	Score 344;	DB 4;	Length 603;
Best Local Similarity	23.38;	Pred. NO. 4.5e-23;		
Matches 141; Conservative	63;	Mismatches 194;	Indels 208;	Gaps 16

[illegible]

Qy	415	EMTDC	420
		11	
Db	599	HfV-HC	603

```

12
US-08-473-089-50
Sequence 50, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Roni, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-0763
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 603 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor bind.
INDIVIDUAL ISOLATE: pro. complex-rat, Fig. 33
US-08-473-089-50

```

Query Match	9.2%	Score 344;	DB 4;	Length 603;
Best Local Similarity	23.3%	Pred. No. 4.5e-23;		
Matches 141;	Conservative	63;	Mismatches 194;	Indels 208;
			Gaps	16;

[illegible]

```

RESULT 13
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Rbdo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2

```

[illegible]

```

0Y 205 -----LDNN 208
Db 255 FVCSDEEGHQSFMAPSCSYLHCPACTCSNNIYDCRGKLTETPTNLPEITTEIRLEON
0Y 209 -----FRPLNLSVLVLAGNULREISDYALEGHQSLESLSFYDNOLAVRR-----
Db 315 TIKYVPGAFSPYKKLRIRIDLSNNOJISELAPDAGFGRSLNSVLIVGNKTKTELPKSLFES
0Y 256 -----ALEQVPGIKFLDLNKNPLQVARGQDPAW-----
Db 375 LFSQLLLLNANKINCLRVDAFODLHNLNLSTLDYDKLQTIAGQTESPLRAIOTMHLAON
0Y 285 -----LHK-----
Db 435 PFICDCHLKMLADYLHTNPITETSGARCTSPRLRANKRIGQTSKSKFPCSTEDYRBSKLS
0Y 289 -----ELGNMMEBVSIDKFAVLNLP
Db 495 DCFADLACPEKRCRGCTVDCSNQKNLKIRPEHIDYAEALRLNNEPTVLEATGIEFKPL
0Y 311 ELTKLDTITNNP-----RLSFHPRAFHNLPOKETTMINNAL
Db 555 QARKINFNSNNKTTIEEGAFEGASGVNEILLTSNRLEVONQHKMKGLESJKTTLKLSNRI
0Y 348 SALHOOT-----VESLPLNLOEYGLHNPTRCDVYIRMANAT
Db 615 TCVGNDSPFGLGSVRLSLVLYNOITTVAPAGAFDPLHLSLSTLNLANPNCNYLAWLGEW
0Y 384 GTRVFFI--EPOSTICAEPPDLQRLPVAEYFFRENT-----DHCLPL 423
Db 675 LKKKILVGNPR--CQKPYFLKEIPIODVAIDQFTDGDGDDSCSPL 720

```

```

?      RESULT 14
? US-09-540-245A-2
? Sequence 2, Application US/09540245A
? Patent No. 6270984
? GENERAL INFORMATION:
? APPLICANT: Goodman, Corey
? APPLICANT: Kid, Thomas
? APPLICANT: Brose, Katja
? APPLICANT: Tessier-Lavigne, Marc
? TITLE OF INVENTION: Modulating Robo: Ligand Interactions
? FILE REFERENCE: B98-031-3
? CURRENT APPLICATION NUMBER: US/09/540,245A
? CURRENT FILING DATE: 2000-03-31
? PRIOR APPLICATION NUMBER: 60/065,544
? PRIOR FILING DATE: 1997-11-14
? PRIOR APPLICATION NUMBER: 60/081,057
? PRIOR FILING DATE: 1998-04-07
? NUMBER OF SEQ ID NOS: 20
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 2
? LENGTH: 1525
? TYPE: PRT
? ORGANISM: human
US-09-540-245A-2
```

[illegible]

Sat Aug 21 14:55:21 2002

us-09-905-088a-245.rai

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:34:58 ; Search time 35.32 Seconds

(without alignments)
2242.231 Million cell updates/sec

Title: US-09-905-088a-245

Perfect score: 3732
Sequence: 1 MRLVAPLLPLAWVAGATATV.....RKLPRSEGETLLPLSQNS 713

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- A.Geneseq_032802:*
- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
 - 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
 - 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
 - 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
 - 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
 - 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
 - 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
 - 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
 - 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
 - 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
 - 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
 - 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
 - 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
 - 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
 - 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
 - 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
 - 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
 - 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
 - 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
 - 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
 - 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
 - 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3732	100.0	713	20	AAV13385
2	3732	100.0	713	22	AAAB80253
3	1993	53.4	377	22	AAAB92765
4	1695	45.4	716	21	AAAB32472
5	1695	45.4	716	22	AAU12427
6	1695	45.4	716	22	AAAB27234
7	1695	45.4	716	22	AAAB50965
8	1695	45.4	716	22	AAAG67530
9	1672	44.8	716	22	AAAM40376
10	1656.5	44.4	705	21	AAAB42513
11	1656.5	44.4	705	22	AAAW8823

12	1656.5	44.4	708	20	AAV13355	Amino acid sequenc
13	1656.5	44.4	708	21	AAV70030	Human extracellular
14	1656.5	44.4	708	22	AAAB39309	Human polypeptide
15	1656.5	44.4	708	22	AAU12320	Human PRO220 polyp
16	1656.5	44.4	708	22	AAAB80223	Human PRO220 prote
17	1654.5	44.3	719	22	AAAT79807	Human protein SRQ
18	1654.5	44.3	719	22	AAAM41095	Human polypeptide
19	1640	43.9	707	20	AAV02379	Human secreted pro
20	1504	40.3	222	22	AAE03524	Human polypeptide
21	875.5	23.5	431	22	AAAB94521	Human secreted pro
22	755.5	20.2	273	22	AAAB93523	Human protein sequ
23	546.5	14.6	592	22	AAE09437	Human sbdJangov9a
24	492.5	13.2	620	22	AAE74705	Human membrane ass
25	490.5	13.1	614	20	AAAB4596	Amino acid sequenc
26	489.5	13.1	620	20	AAV13357	Amino acid sequenc
27	489.5	13.1	620	22	AAU12333	Human PRO227 polyp
28	489.5	13.1	620	22	AAAB80225	Human PRO227 prote
29	468.5	12.6	606	22	AAE06799	Human neuronal gui
30	467.5	12.5	579	22	AAE06804	Mature human neuro
31	442	11.8	1091	18	AAWA1641	Sequence used in d
32	442	11.8	1091	20	AAV08099	Murine glial cell
33	442	11.8	1091	20	AAV08010	Mouse LIG-1 protei
34	442	11.8	1091	21	AAE97833	Murine LIG-1 prote
35	438.5	11.7	640	20	AAV08100	Human PRO331 prote
36	438.5	11.7	640	20	AAAB5722	Novel protein (C1o
37	438.5	11.7	640	20	AAV13394	Amino acid sequenc
38	438.5	11.7	640	21	AAAB24407	Human PRO331 prote
39	438.5	11.7	640	21	AAV70673	Human PRO331 prote
40	438.5	11.7	640	22	AAU12355	Human PRO331 polyp
41	438.5	11.7	640	22	AAV00826	Human immune respo
42	438.5	11.7	640	22	AAAB80262	Human PRO331 prote
43	438.5	11.7	640	22	AAAB5292	Human PRO331 prote
44	438.5	11.7	640	22	AAAB53089	Human angiogenesis
45	432.5	11.6	540	22	AAAB5612	Drosophila melanog

ALIGNMENTS

RESULT 1	
ID	AAV13385
AC	AAV13385 standard; Protein: 713 AA.
XX	
XX	AAV13385;
XX	
DT	25-JUN-1999 (first entry)
XX	
XX	
DE	Amino acid sequence of protein PRO293.
XX	
KW	Secreted protein; transmembrane protein; human; enterocolitis;
KW	Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW	congenital microvillus atrophy; skin disease; cell growth;
KW	abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW	Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW	fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW	anti-thrombotic; wound healing; tissue repair.
XX	
OS	Homo sapiens.
XX	
PN	WO9914328-A2.
PD	25-MAR-1999.
XX	
XX	
PF	16-SEP-1998; 98WO-US19330.
XX	
XX	25-NOV-1997; 97US-0066840.
PR	17-SEP-1997; 97US-0059113.
PR	17-SEP-1997; 97US-0059115.
PR	17-SEP-1997; 97US-0059117.
PR	17-SEP-1997; 97US-0059119.
PR	17-SEP-1997; 97US-0059121.
PR	17-SEP-1997; 97US-0059122.
PR	17-SEP-1997; 97US-0059184.

CC	PR	18-SEP-1997;	97US-0059263.
CC	PR	18-SEP-1997;	97US-0059266.
CC	PR	15-OCT-1997;	97US-0062125.
CC	PR	17-OCT-1997;	97US-0062285.
CC	PR	17-OCT-1997;	97US-0062287.
CC	PR	21-OCT-1997;	97US-0063486.
CC	PR	24-OCT-1997;	97US-0062814.
CC	PR	24-OCT-1997;	97US-0062816.
CC	PR	24-OCT-1997;	97US-0063045.
CC	PR	24-OCT-1997;	97US-0063120.
CC	PR	24-OCT-1997;	97US-0063121.
CC	PR	24-OCT-1997;	97US-0063127.
CC	PR	24-OCT-1997;	97US-0063128.
CC	PR	27-OCT-1997;	97US-0063329.
CC	PR	27-OCT-1997;	97US-0063327.
CC	PR	28-OCT-1997;	97US-0063541.
CC	PR	28-OCT-1997;	97US-0063542.
CC	PR	28-OCT-1997;	97US-0063544.
CC	PR	28-OCT-1997;	97US-0063549.
CC	PR	28-OCT-1997;	97US-0063550.
CC	PR	28-OCT-1997;	97US-0063564.
CC	PR	29-OCT-1997;	97US-0063435.
CC	PR	29-OCT-1997;	97US-0063704.
CC	PR	29-OCT-1997;	97US-0063732.
CC	PR	29-OCT-1997;	97US-0063738.
CC	PR	29-OCT-1997;	97US-0063734.
CC	PR	29-OCT-1997;	97US-0064215.
CC	PR	29-OCT-1997;	97US-0063735.
CC	PR	31-OCT-1997;	97US-0063870.
CC	PR	31-OCT-1997;	97US-0064103.
CC	PR	03-NOV-1997;	97US-0064248.
CC	PR	07-NOV-1997;	97US-0064809.
CC	PR	12-NOV-1997;	97US-0065186.
CC	PR	17-NOV-1997;	97US-0065846.
CC	PR	18-NOV-1997;	97US-0065693.
CC	PR	21-NOV-1997;	97US-0066120.
CC	PR	21-NOV-1997;	97US-0066364.
CC	PR	24-NOV-1997;	97US-0066772.
CC	PR	24-NOV-1997;	97US-0066466.
CC	PR	24-NOV-1997;	97US-0066770.
CC	PR	24-NOV-1997;	97US-0066511.
CC	PR	24-NOV-1997;	97US-0066453.
XX	PA	(GETH) GENENTECH INC.	
XX	PI	Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;	
XX	DR	WPI: 1999-229533/19.	
XX	DR	N-PSDB: AAX52256.	
XX	PT	New isolated human genes and polypeptides used in, e.g. treatment of	
XX	PS	gastrointestinal ulceration	
XX	XX	Claim 12; Fig 86; 320pp; English.	
CC	CC	AAV1344-403 represent secreted and transmembrane human proteins.	
CC	CC	The cDNA sequences are obtained from cDNA libraries, prepared from	
CC	CC	fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.	
CC	CC	The encoded polypeptides have specific uses based on their homology to	
CC	CC	known polypeptides, e.g. PRO211 and PRO217 can be used for disorders	
CC	CC	associated with the preservation and maintenance of gastrointestinal	
CC	CC	mucosa and the repair of acute and chronic mucosal lesions	
CC	CC	(e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal	
CC	CC	ulceration and congenital microvillus atrophy), skin diseases associated	
CC	CC	with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial	
CC	CC	cancers such as lung squamous cell carcinoma of the vulva and gliomas),	
CC	CC	potent effects on cell growth and development, diseases related to	
CC	CC	growth or survival of nerve cells including Parkinson's disease,	
CC	CC	Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as	
CC	CC	for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used	
CC	CC	as a target for anti-tumor drugs. PRO533 may be used in the treatment	
CC	CC	of Usher Syndrome or Atrophic areata; PRO269 can be used as an	
CC	CC	anti-thrombotic agent; PRO287 polypeptides and portions may have	

Query Match	100.0%	Score 3732	DB 20	Length 713
Best Local Similarity	100.0%	Pred. No. 1e-290		
Matches 713	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	MLLVAPLLILAMVAGATATVPVYVPHVHVCPRQCAQIRPWTPRPSRYEATVDCNDLFL	60		
DB 1	MLLVAPLLILAMVAGATATVPVYVPHVHVCPRQCAQIRPWTPRPSRYEATVDCNDLFL	60		
QY 61	TAVPALPAGTQTLTLQSSNIVYVQSEGLYANTTELDSLQNSPSDARDDEPHALPOLL	120		
DB 61	TAVPALPAGTQTLTLQSSNIVYVQSEGLYANTTELDSLQNSPSDARDDEPHALPOLL	120		
QY 121	SLHLEENQITRLLEDHSFAGLASLOELVLDNHQLYRIAPRAFSGLSNLRLLNSNLRAI	180		
DB 121	SLHLEENQITRLLEDHSFAGLASLOELVLDNHQLYRIAPRAFSGLSNLRLLNSNLRAI	180		
QY 181	DSRWEMLPNLEITLIGKNKYDAIILDAMPRLANIRSLVLAGMNIRESIDYALBSLGLE	240		
DB 181	DSRWEMLPNLEITLIGKNKYDAIILDAMPRLANIRSLVLAGMNIRESIDYALBSLGLE	240		
QY 241	SLSPFDNQLARPRRALQOVPGIKFLDINKPKPLQGVGSDPANMHLKELGINNNEELVS	300		
DB 241	SLSPFDNQLARPRRALQOVPGIKFLDINKPKPLQGVGSDPANMHLKELGINNNEELVS	300		
QY 301	IDKFAVLNLPETFKLDITNNPRLFIDBRATHHLPQMETLMLNNNALSALHQGVESLPN	360		
DB 301	IDKFAVLNLPETFKLDITNNPRLFIDBRATHHLPQMETLMLNNNALSALHQGVESLPN	360		
QY 361	LQVGLHGNPIRCCOVIMANATGRRVFTIERQSTLCAPRPDLQRLPYREVPFRMTDHC	420		
DB 361	LQVGLHGNPIRCCOVIMANATGRRVFTIERQSTLCAPRPDLQRLPYREVPFRMTDHC	420		
QY 421	LPILSPRSFSPSLQVASESMVILHCRALAEPEPELTYWTYPAGLRLTPAHAGRRRVYPEG	480		
DB 421	LPILSPRSFSPSLQVASESMVILHCRALAEPEPELTYWTYPAGLRLTPAHAGRRRVYPEG	480		
QY 481	TLELRVTAEEAGLYCYAQNILVGADRTKTVSVVYGRALLQFGRDGQGLELRYOETBRYH	540		
DB 481	TLELRVTAEEAGLYCYAQNILVGADRTKTVSVVYGRALLQFGRDGQGLELRYOETBRYH	540		
QY 541	ILLSWVTPPNVYVSTNLTSSASSLRGOCATALARLPKRTHSYNTIRLLQATETYNACLOYA	600		
DB 541	ILLSWVTPPNVYVSTNLTSSASSLRGOCATALARLPKRTHSYNTIRLLQATETYNACLOYA	600		
QY 601	FADAHQTOACWAWARKKEKTSCHRALGDPGGLIATILALVLLAAGLAHLTGGPRKGVG	660		
DB 601	FADAHQTOACWAWARKKEKTSCHRALGDPGGLIATILALVLLAAGLAHLTGGPRKGVG	660		
QY 661	GRRPLPPAMAFWGSAPSVRVYSAPLVLPWNPGRKLPRSSGEETLLPPLSQNS 713			
DB 661	GRRPLPPAMAFWGSAPSVRVYSAPLVLPWNPGRKLPRSSGEETLLPPLSQNS 713			
RESULT 2				
AAAB80253	standard; Protein: 713 AA.			
AAAB80253				
24-APR-2001	(first entry)			
Human PRO293	protein.			

KW	antiatherogenic, antiinfertility; antidiabetic; antiviral; diabetes;
KW	ophthalmologic; gene therapy; skin disease; gastrointestinal disorder
KW	ischemia; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	WO200104311-A1.
XX	
PD	18-JAN-2001.
XX	
PF	22-FEB-2000; 2000WO-US04414.
XX	
PR	07-JUL-1999; 99US-0143048.
PR	26-JUL-1999; 99US-0145698.
PR	28-JUL-1999; 99US-0146222.
PR	08-SEP-1999; 99WO-US20254.
PR	13-SEP-1999; 99WO-US202944.
PR	15-SEP-1999; 99WO-US21090.
PR	15-SEP-1999; 99WO-US21547.
PR	05-OCT-1999; 99WO-US23089.
PR	29-NOV-1999; 99WO-US28214.
PR	30-NOV-1999; 99WO-US28313.
PR	16-DEC-1999; 99WO-US30095.
PR	20-DEC-1999; 99WO-US30911.
PR	20-DEC-1999; 99WO-US30999.
PR	05-JAN-2000; 99WO-US00219.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Ashkenazi AJ, Bolstein D, Desnoyers L, Eaton DL, Ferrara N;
PI	Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI	Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavlin IJ;
PI	Mathar JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI	Williams PM, Wood WI;
XX	
DR	WPI, 2001-081051/09.
XX	
XX	N-PSSDB; AAF72414.
PT	sixty one nucleic acids encoding PRO polypeptides which are useful in
PT	the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT	squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT	Alzheimer's disease) -
XX	
XX	Claim 1, Fig 86; 393pp: English.
XX	
CC	The present sequence is one of sixty one novel secreted and
CC	transmembrane PRO polypeptides. The PRO polypeptides are
CC	useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC	squamous cell carcinoma), gastrointestinal disorders (e.g. lung
CC	enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC	Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC	peripheral bleeding angiogenesis, ischemias such as coronary
CC	ischemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC	rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC	diabetes and retinal disorders such as retinitis pigmentosa.
CC	The PRO nucleic acids have applications in molecular biology, including
CC	use as hybridization probes, and in chromosome and gene mapping.
XX	
XX	Sequence 713 AA:

Query Match	100.0%;	Score 3732;	DB 22;	Length 713;
Best Local Similarity	100.0%;	Pred. No. 1e-290;		
Matches 713; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

Qy	1	MLRLVAPLLLA	VAAATATVTVVPMWHPGCPDCAACIRPMYPRSSRYRATVPOCNDL	60
Db	1	MLRLVAPLLLA	VAAATATVTVVPMWHPGCPDCAACIRPMYPRSSRYRATVPOCNDL	60
Qy	1	MLRLVAPLLLA	VAAATATVTVVPMWHPGCPDCAACIRPMYPRSSRYRATVPOCNDL	60
Db	1	MLRLVAPLLLA	VAAATATVTVVPMWHPGCPDCAACIRPMYPRSSRYRATVPOCNDL	60
Qy	61	TAVPALPAGT	OTLLQSNSTIVRDSGLYLANITLDELSONSFSDARDCDFALPOL	120
Db	61	TAVPALPAGT	OTLLQSNSTIVRDSGLYLANITLDELSONSFSDARDCDFALPOL	120

Qy	121	SLLEENQUTLREDEHSFAGLASLOEYKLNNOQYXRIMPRAFSGSNLTRHLSNNILRA	180
Db	121	shlleengtlrtledhstnglaslqelylnhnglylrslaprtsglsnlltrhlsnnlra	180
Qy	181	DSRMEEMPRNEIIMIGKNKVDALILDMNFRPLANLSVLAGMULREISYALEGLOSLE	240
Db	181	dsrtemlprnelimignkvdallmfnplnlltrslvagmnlreisyaleglsle	240
Qy	241	SLSFYDNOLAVPRALOEYVGLKFLDLNKNPLQRVGPDPANMLHUKELGNNMEELVS	300
Db	241	slsfydnqlavpralaeqvgjklfldlnknplrqrvpgdfammlhukelgnmeelvs	300
Qy	301	IDKRALVNLPELTKLDTNNRSLSEFHPRANHILPQWETMLNNAALSHOQTVESTLPN	360
Db	301	ldkrlavnlpeltkldtlnnprrslfhprrahhlpqmetlmlnnaalsahqvtvestlpn	360
Qy	361	LOEVLGHONPRCOCVIRWMANATRYAFIEPOSTLCAEPPDLORLVREVPREEMDHC	420
Db	361	lgevlglnprrcdcvirwanaatryrfliepgstlcaepdpdqlrvreyrfremtdhc	420
Qy	421	LPLISPRSPSLOVASGESMWLHCRALEBEPRIYWTAPAGRLPRAHAGRRYRYPEG	480
Db	421	lplisprsfpslqvasgesmwlhcralaebepriywtapagrlprahagrryrypeg	480
Qy	481	TLELRRTVAEEAGLYTCAQNLVADNRKTVSVVVGRRLLQVRGREGGELTRQETHRPN	540
Db	481	tlelrvtvaeaaglytcaqnlvadrktvsvvvgrrllpvydreggslrlyrgethrpn	540
Qy	541	ILLSWTPRPNTVSNFLTWSSASLSRGCGATLALPRGTHSYNTRTLQLOTEWACLOVA	600
Db	541	lllswtprpntvsnfltwssaslsrggatabalprgthsyntrtlqgteywaclova	600
Qy	601	FADAHQTOLACYWARTKEATSCHRALGDRPGILAILALAVLLAAGLAHLGTGQPRKGVG	660
Db	601	fadahtqlacwartkeatschralgdrpgilailavlllaaglaahlgtygprkvg	660
Qy	661	GRRPLPRMAWGMGSAFSVRYSAPLYLRPMWPGKLRSSGCELTLPRLPSLONS 713	
Db	661	grrplprmafwgmsafsvryvasplylrpmwpgkrlrpsgeceltrpplsqs 713	

RESULT	3
AAB92765	
ID	AAB92765 standard; Protein: 377 AA.
XX	
AC	AAB92765;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:11234.
XX	
KW	Human; primer: detection; diagnosis; antisense therapy; gene therapy
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PE	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isegaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
XX	Ishtai S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR	WPI, 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS
PS Claim 8; SEQ ID 11234; 2537pp + CD ROM; English.
XX
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX
XX Sequence 377 AA;

	Query Match	53.4%	Score 1993;	DB 22;	Length 377;
	Best Local Similarity	100.0%	Pred. No. 1.0e-151;		
	Matches 377;	Conservative	0;	Mismatches 0;	Indels 0;
					Gaps 0;
QY	337	METLMLNNNALSLALQCOVVESELPNIOEVLGHANPNIRPCDCVIRMANATGRRVFIEPOSTL	396		
Db	1	metlmlnnnalsalnlgqvleslpmngvlglnpnrlrcdcvirwanatgrrvffiepgstl	60		
QY	337	CAEPDLORLPYREVPFRBMDHCLPLISPSRSPSLOVASGESMWLHCRALAEPEPEY	456		
Db	61	caepdqlrlpyreyefremthcncplispsrfsplqvasgesmwlnhcralaepepeyl	120		
QY	457	WTPRGLRFLTRPHACRRRVYRPEGFLTELIRVYAEAEAGLYTCANOLNIGADRTTVEVYGR	516		
Db	121	wtprgldrltrphagrtryrvypegclteltrvyaeeaglytcvaqnlvgadkltvsvvygr	180		
QY	517	ALLQGRBEGGGLERVRVETHPYHLLSMWTPRPNVSTNLTWSSASSLSRGOCATLARLP	576		
Db	181	allqgridegggltrvgethpyhlllswtrpctvstnltwssasslrrggataalarlp	240		
QY	577	RGTHSYNTRILLQATNEYAACLOVAPADAHQTOLACWARTKEATSCSHRALGDRPGLIATIA	636		
Db	241	rgthsyntrilllqatecyacilqvatadahtqlacwarktkeatschrlgdrpjlialia	300		
QY	637	LAVLLLAAGLAHAHLTGQPRKGVGGRARPLPMAWMAWGSAPSVRVVSAPLVYPMWNGRTL	696		
Db	301	lavlllaaglaahhlgvgprkvgvggrlppawatwgsapsrvrvvseplvlvpmwngrtl	360		
QY	697	PRSEGEFTLLPPLSONS	713		
Db	361	prsegecllplpsqns	377		

RESULT	4
AAB33472	
ID	AAB33472 standard; Protein; 716 AA
XX	
AC	AAB33472;
XX	

DT	29-JAN-2001	(first entry)
XX		
DE	Human	PRO1338 protein UNO693 SEQ ID NO:279.
XX		
KW	Human; immune related disease; diagnosis; antiinflammatory; cardiant;	
KW	dermatologic; anitlathritic; antirheumatic; immunosuppressive;	
KW	haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;	
KW	antianaemic; hepatologic; virtude; antipsoriatic; antiallergic;	
KW	antiaesthatic; systemic lupus erythematosus; rheumatoid arthritis;	
KW	osteoarthritis; spondyloarthopathy; systemic sclerosis; sarcoidosis;	
KW	idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;	
KW	systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;	
KW	autoimmune thrombocytopaenia; immune-mediated renal disease;	
KW	demylinating disease; hepatobiliary disease; Multiple's disease;	
KW	inflammatory bowel disease; gluten-sensitive enteropathy;	
KW	autoimmune disease; immune-mediated skin disease; allergic disease;	
KW	immunological disease; transplantation associated disease;	
KW	graft rejection; graft-versus-host-disease.	
OS		
XX	Homo sapiens.	
PN		
XX	WO200053758-A2.	
PD	14-SEP-2000.	
XX		
PF	02-MAR-2000; 2000WO-US05841.	
XX		
PR	08-MAR-1999; 99WO-US05028.	
PR	10-MAR-1999; 99US-0123618.	
PR	12-MAR-1999; 99US-0123957.	
PR	13-MAR-1999; 99US-0125775.	
PR	12-APR-1999; 99US-0128849.	
PR	20-APR-1999; 99WO-US08615.	
PR	28-APR-1999; 99US-0131445.	
PR	04-MAY-1999; 99US-0132371.	
PR	14-MAY-1999; 99US-0134287.	
PR	02-JUN-1999; 99WO-US12252.	
PR	23-JUN-1999; 99US-0141037.	
PR	20-JUL-1999; 99US-0144758.	
PR	26-JUL-1999; 99US-0145698.	
PR	28-JUL-1999; 99US-0146222.	
PR	01-SEP-1999; 99WO-US20111.	
PR	08-SEP-1999; 99WO-US20594.	
PR	13-SEP-1999; 99WO-US20944.	
PR	15-SEP-1999; 99WO-US21090.	
PR	15-SEP-1999; 99WO-US21547.	
PR	05-OCT-1999; 99WO-US23089.	
PR	29-OCT-1999; 99US-0162506.	
PR	29-NOV-1999; 99WO-US28214.	
PR	30-NOV-1999; 99WO-US28313.	
PR	30-NOV-1999; 99WO-US28409.	
PR	01-DEC-1999; 99WO-US28301.	
PR	01-DEC-1999; 99WO-US28654.	
PR	02-DEC-1999; 99WO-US28551.	
PR	02-DEC-1999; 99WO-US28564.	
PR	02-DEC-1999; 99WO-US28565.	
PR	16-DEC-1999; 99WO-US30095.	
PR	20-DEC-1999; 99WO-US30999.	
PR	30-DEC-1999; 99WO-US31274.	
PR	05-JAN-2000; 2000WO-US00219.	
PR	06-JAN-2000; 2000WO-US00277.	
PR	06-JAN-2000; 2000WO-US00376.	
PR	11-FEB-2000; 2000WO-US03565.	
PR	18-FEB-2000; 2000WO-US04341.	
PR	18-FEB-2000; 2000WO-US04342.	
PR	22-FEB-2000; 2000WO-US04414.	
XX		
PA	(GETH) GENENTECH INC.	
PI	Ashtkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W,	
XX	Kaakakoff RC, Lu Y, Pan J, Pennica D, Shelton DU, Smith V,	
PI	Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;	
XX		


```
Db      210  |||:|||||:|||||:|:| ||| |||:|||||:|:|:|:|:| ||| ||| |||
210  nfkplaniatsivlagmyltdipgnalvgldslsleslftdnlklykvpqlaigvynlkfId 269
QY      268  LKNKPLQVRGPGDFANMLHLKEGLNNMEELVSDKFAVLNDELTKLDTNNPRLSFTN 327
Db      270  lknkplhkiqgdtkmnlrklkeglngmngelvsdvayalnlpelctkleatnprklsyln 329
QY      338  PRAFHNLPMQETLMNNALSAHQOYVESLPNLQEVGLGNPRICDCVIRMANATGTRY 387
Db      330  rlafrsvpaleslmnnalnaikyktvesipnlreislinsprlrcdcvlnhwnsktnl 389
QY      388  RPIEPQSTLCAEPDLOQLPYREVPFRMDHCPLTSPSPFSLOVASGEEMVLHCRA 447
Db      390  rtmepismfcampyekhgykveylldgsseqclpmshdsfporlnvdigtvtfidcra 449
QY      448  LAEPPELYWYTPAGRLRTPAHAGRRYRVPEGTLELRVTAEEAGLYTCVAONLVGADT 507
Db      450  maepelywlytprlgnkltvetltsdkyklsesgclslnsqldsgsrgytcvaqnvqadt 509
QY      508  KTVSVVGRALLQPGRDGQGLELKVQETHHYHLLSWVTPPNVSTNLTWSSAS-SLRG 566
Db      510  rvaclkvngtll---dglgylkyvkqteshslvswkynsnvmstnlkwsatmkIdn 565
QY      567  OGATLALRLPRGTHSYNTRILQATEYWAQLQVAFADAHQTOLACVWARTKEA----- 618
Db      566  plitylarvpydvheynlthlqptdyevcltvsnlhgtqkscvntknaafavdlad 625
QY      619  ----TSCRALGDRPGLAIITLALAV 639
Db      626  getstalaavmgsmfavistlaslav 650
```

```
RESULT 7
AAB50965
ID      AAB50965 standard; Protein: 716 AA.
XX
AC      AAB50965;
XX
DT      21-MAR-2001 (first entry)
XX
DE      Human PRO1338 protein.
XX
KW      Human; PRO; cytosolic; nootropic; neuroprotective; respiratory general;
KW      antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;
KW      PRO agonist; cancer; inflammatory disorder; immunological disorder.
XX
OS      Homo sapiens.
XX
PN      WO200073348-A2.
PD      07-DEC-2000.
PF      30-MAY-2000; 2000WO-US14941.
XX
PR      02-JUN-1999; 99WO-US12252.
PR      22-JUN-1999; 99US-0140650.
PR      23-JUN-1999; 98US-0144758.
PR      20-JUL-1999; 98US-0144758.
PR      01-SEP-1999; 99WO-US20111.
PR      08-SEP-1999; 99WO-US20594.
PR      29-OCT-1999; 99US-0162506.
PR      30-NOV-1999; 99WO-US28313.
PR      01-DEC-1999; 99WO-US28634.
PR      02-DEC-1999; 99WO-US28651.
PR      16-DEC-1999; 99WO-US30095.
PR      20-DEC-1999; 99WO-US30999.
PR      06-JAN-2000; 2000WO-US00376.
PR      11-FEB-2000; 2000WO-US03565.
PR      18-FEB-2000; 2000WO-US04341.
PR      18-FEB-2000; 2000WO-US04342.
PR      02-MAR-2000; 2000WO-US05841.
PR      03-MAR-2000; 2000US-0187202.
PR      10-MAR-2000; 2000WO-US06319.
```

```
PR      15-MAR-2000; 2000WO-US06884.
PR      30-MAR-2000; 2000WO-US08439.
PR      17-MAY-2000; 2000WO-US13705.
XX
PA      (GETH ) GENENTECH INC.
PI      Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;
PI      Shelton DL, Smith V, Watanabe CK, Wood WI;
DR      WPI; 2001-016509/02.
DR      N-PSDB; AAC91567.
XX
XX      Twenty eight nucleic acids encoding PRO polypeptides which are useful
XX      for treating various tumors, e.g. breast cancer, and other
XX      inflammatory, angiogenic and immunological disorders -
XX
XX      Claim 31; Fig 30; 188pp; English.
XX
CC      The present sequence is one of twenty eight novel PRO polypeptides. The
CC      PRO polypeptides and their agonists, including antibodies, peptides, and
CC      small molecule agonists, may be used to treat various tumors, e.g.,
CC      cancers such as breast cancer, ovarian cancer, renal cancer, colorectal
CC      cancer, uterine cancer, prostate cancer, lung cancer, bladder cancer,
CC      central nervous system cancer, melanoma or leukaemia. They are also
CC      useful for treating other disorders such as neuronal, glial, astrocytal,
CC      hypothalamic and other glandular, macrophagal, epithelial, stromal and
CC      blastocoelec disorders, and inflammatory, angiogenic and immunological
CC      disorders.
XX
XX      Sequence 716 AA;
SQ
```

```
Query Match 45.4%; Score 1695; DB 22; Length 716;
Best Local Similarity 53.6%; Pred. No.3.5e-127;
Matches 335; Conservative 104; Mismatches 166; Indels 20; Gaps 5;

QY      29  CPPOCAGQIRPWYTPRSSRYRATVDCNDLFLAVPALPAGTQTLLOSNSIVR-VDGS 87
Db      32  cpqlcevcelrpfwpfgytreatvdcndlrtrtrpnlssdcvylldgmnaktvcd-- 89
QY      88  ELGYLANLTLELDLSONSFSDARDCDFHALPOLSLHLEONOLRLEHSPAGLSLOELY 147
Db      90  elqqlfnlteldfsgnlnlkevglanlqtlthleeqtlemtycldslsly 149
QY      148  LNNHOLYIAPRASGSLNLRHLNLSLRAIDSRFEMLPNLEITIMIGKNRVADILDM 207
Db      150  lnhqstlsahafaglnlrlhlnsnklkldsrwfdstpnleilmigenpvgldm 209
QY      208  MPRPLANRSLVLAGMNLREISDYALBSLOSLESLSPYDQOLARVPRALQVPGKFLD 267
Db      210  nfkplaniatsivlagmyltdipgnalvgldslsleslftdnlklykvpqlaigvynlkfId 269
QY      268  LKNKPLQVRGPGDFANMLHLKEGLNNMEELVSDKFAVLNDELTKLDTNNPRLSFTN 327
Db      270  lknkplhkiqgdtkmnlrklkeglngmngelvsdvayalnlpelctkleatnprklsyln 329
QY      328  PRAFHNLPMQETLMNNALSAHQOYVESLPNLQEVGLGNPRICDCVIRMANATGTRY 387
Db      330  rlafrsvpaleslmnnalnaikyktvesipnlreislinsprlrcdcvlnhwnsktnl 389
QY      388  RPIEPQSTLCAEPDLOQLPYREVPFRMDHCPLTSPSPFSLOVASGEEMVLHCRA 447
Db      390  rtmepismfcampyekhgykveylldgsseqclpmshdsfporlnvdigtvtfidcra 449
QY      448  LAEPPELYWYTPAGRLRTPAHAGRRYRVPEGTLELRVTAEEAGLYTCVAONLVGADT 507
Db      450  maepelywlytprlgnkltvetltsdkyklsesgclslnsqldsgsrgytcvaqnvqadt 509
QY      508  KTVSVVGRALLQPGRDGQGLELKVQETHHYHLLSWVTPPNVSTNLTWSSAS-SLRG 566
Db      510  rvaclkvngtll---dglgylkyvkqteshslvswkynsnvmstnlkwsatmkIdn 565
QY      567  OGATLALRLPRGTHSYNTRILQATEYWAQLQVAFADAHQTOLACVWARTKEA----- 618
```

1

PN W0200153312-A1.

W0200153312-A1

PD 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
PF
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI59532.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
PS Example 6: SEQ ID NO 3521: 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAI38642-AAI42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
CC
CC
XX Sequence 716 AA;
SQ

Query Match 44.8%; Score 1672; DB 22; Length 716;
Best Local Similarity 53.1%; Pred. No. 2.5e-125;
Matches 332; Conservative 105; Mismatches 168; Indels 20; Gaps 5;

QY 29 CPGCACQIRPWYPRSSYREATVDCNDLFLAVPALPAGTQTLLQSNSTIVR-VDQS 87
DB 32 cpqlcveclrpwftpgstyreattvdcndlrtrtrpnsldsgvlllqsnlaktvd-- 89
QY 88 ELGTIANITELDLQNSFSQARDCDFHALPOLSLHLENOQLRLRLEDHRSAGIASIOELX 147
DB 90 elqgfnlteldfsqnmftlnkvglanltqtlthleeqqtemtycdqldsnlgely 149
QY 148 LNHOLYRIAPRAFSGSLNLRLLHNSNLRAIDSRFEMIPNLEIMIGKNVVDAILDM 207
DB 150 lnhglsfslshafaglknlrlhlnsnkikvidrstfdstpyldlmisgpnvlgldm 209
QY 208 NFRPLANTRSVLVLAGMNIRESIDYALBGLQSLSESLFYDQNLARVPRRALEOVGKLFD 267
DB 210 nfrplantrsvlvlagmyltdipgnalvglslesltydunklvyvqplaigkypnklfld 269
QY 268 LKNNKLOLVGPGDFANMHLKELGNMNEELVSDKFAIVNLDELTLDTNNPRLSEFH 327
DB 270 lknknlolvpgdfkmmrlrkeigimngelvsydrtaidnlpeltkleestmpklsylh 329
QY 328 PRAFNHLRPMETLMLNNALSALHQQTVESLPNLQEVGLGNPLRCDCVIRMANATCTRY 387
DB 332 prafnhlrmetlmlnnalsalhqqtveslpnlqevglgnplrcdcvirmanatctry 387

DB 330 rlafrswpaleslmlnnalnaityktveslpnlreislsbnplrcdewihtrnsknki 389
QY 388 RFIPEOSTICAEPPDLQRLPYREVPFREMTHDCLPLTSPRSFPSPQVSGESMVJHCRA 447
DB 390 rfimeplmfcmpepyghykvkelyldqsdsegclpmashsfprlnvtdigtvtfldcra 449
QY 448 LAEPPEIYWTYPAGLNLTPAHAGRRYRVPESTLEIRRYTAEEAGLYTCVAQNLVGADT 507
DB 450 maeppeiywtypglnltvetlsdkylissegtleisnigtledsgrylcvaqnvgnadt 509
QY 508 KTVSVYVGRALLQGRDEGGLELRQETHPHYHLLSMWTPPTVSTNLTWSSAS-SLHG 566
DB 510 rvaclkvngcll----qgtqylkyvqcteshslvskvsnvmcsnlkwsaakmklon 565
QY 567 QGATALARKPRGTHSYNITRLQATEYMACLOVAFDAHTOLACVWARKKA----- 618
DB 566 phlytarvpdvdyheynlthlqpsdyevclsvsnlhqgqkscvntcknaafavdiad 625
QY 619 ----TSCRHALGDRPGLAIIALAV 639
DB 626 qetstalaavngsmfavislastav 650

RESULT 10
AAB42513
ID AAB42513 standard; Protein: 705 AA.
XX
XX AAB42513;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2277 polypeptide sequence SPQ ID NO:4554.
DE
XX
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerrary; antipsoiatric; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000WO-US08621.
PE
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Leach M;
PI WPI: 2000-602362/57.
DR N-PSDB; AAC76722.
XX
XX Novel nucleic acid and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PR neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 3752-3754; 5507pp; English.

CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SO Sequence 708 AA:

Query Match 44.4%; Score 1656.5; DB 22; Length 708;

Best Local Similarity 47.2%; Pred. No. 4.3e-124;

Matches 333; Conservative 109; Mismatches 217; Indels 47; Gaps 5;

```
QY 17 TATPYVPMHVPCCPQCACQIRWYTPRSSYRATVDCNDLFLTAVPPALRAGTQTL 76
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17 tllvgavdkkvdcpctctcelpwfpctpslymaestvdcndlgllfparlpantglll 76

QY 77 QSNISIVVDOSLGYLANLTLELDLSQNSFSDARCDFFHALPQLLSLLENOLTRLEDHS 136
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 qtnlaktieys-tcfpynlgldtsqnlssvtnlnvkkmpqllsvyleenkltelpkc 135

QY 137 FAGLASLOELYNOLYRIAPRAFSGLSNLHLNLSNLRAIDSRWFEMPLETMT 196
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 136 lseisnqelyinhnlstspgafiglhnllrhlnsnrlqmlnskwidapnllelml 195

QY 197 GGNKVDAIIDMNFPLANLSLVLAGNLTREISDYALEGLOSLESSEFYDNOLARVPRRA 256
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 196 genplrtikmnpfplnrlstlvaqnlteipdnalvglenleslsfydnrltkvphva 255

QY 257 LEQVPGKFLDLNKNPLQVRGPGDFANMLHKLGLNNMEELVSDKEALVNLPELT 316
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 256 lqkvnalkfildnknprntrirgdfsmllhklgeljnmpeilsidslavdnldlrkie 315

QY 317 ITNPRLSFTTPRAFHLPQMETIMLNNALSLAHQOTVESLPNLOEVGHGMPTRCDCV 376
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 316 atnprlsytlpnaftrlpkleslmnsalsalynqleslplnlkelslnsptrcdcv 375

QY 377 IRMANATGTRVRLIEPOSTCAEBPDQLRLPREVPFEREMTDHCLPLISPSFPSLOYA 436
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 376 lrmmmmktnlrfmepdlsfcvdprefgqgvnyvfhfrdmeicplliapesfplnve 435

QY 437 SGESMVLHCRALAREPEITWVPAGLRLTPAHAGRYRVYRPGCTLELRVTAEEAGLYT 496
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 436 agsyvstfncracaepqelywlpqsgkllpntldkfyvhsegldingvtrpkegglyt 495

QY 497 CVAQNLVGADTKTVSVVVGRLLOPRGDEQGLERVOETHPRPHILDSWTPENTVSTNL 556
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 496 clatnlvgadlksvmlkvdsfpq--dnngslnikirldigansvlyvswkaskllksav 552

QY 557 TWSASSSLRGOGATALARPGRGHSYNTIRLQATEWACLOYAFADAHQOLACVAKTK 616
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 553 kwlatfctenshaagsarlpdvkvynltnlpsteykicldipctlyqknrkxkvntck 612

QY 617 -----EATSCHRALGDRGLAIALAVLLAAGLAHGLGTGPRKGVGRR- 663
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 613 glhpqkeyeknttllmacigllgfyicllscf-----spemncdgyhs 660

QY 664 -----PLPRAMAFWGSAPSVRYVSAPLVLPWN 691
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 661 yvrnylqptfalgelypplnlwagkeksstlkvkatvlgipfn 706
```

Search completed: August 31, 2002, 14:36:09
Job time: 71 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:34:58 ; Search time 24.35 Seconds
(without alignments)
2813.622 Million cell updates/sec

Title: US-09-905-088A-245

Perfect score: 3732
Sequence: 1 MRLLVAPLLAVLWAGATATV.....RKLPSSRGETLLPPLLSONS 7131

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1649.5	44.2	707	2 JC7763	neuronal leucine-r
2	442	11.8	1091	2 A58532	glial cell membran
3	407.5	10.9	605	2 A41915	insulin-like growt
4	397.5	10.7	605	2 JC5239	insulin-like growt
5	381.5	10.2	789	2 T28714	hypothetical prote
6	381.5	10.2	1355	2 T28715	hypothetical prote
7	368	9.9	907	2 UC0193	G protein-coupled
8	360.5	9.7	1531	2 T42218	slit-1 protein hom
9	360	9.6	907	2 J50176	orphan G protein-c
10	346	9.3	1523	2 T13953	MEGF5 protein - ra
11	344	9.2	603	2 JC1282	insulin-like growt
12	342.5	9.2	603	2 JC6128	insulin-like growt
13	324	8.7	1469	2 B36655	slit protein 2 pre
14	324	8.7	1480	2 A36655	slit protein 1 pre
15	321	8.6	560	2 A60164	platelet membrane
16	317.5	8.5	702	2 T21148	hypothetical prote
17	302	8.1	1025	2 T42626	secreted leucine-r
18	299.5	8.0	361	2 A53860	chondroadherin pre
19	298	8.0	536	2 A34901	lysine carboxypept
20	288	8.0	680	2 T19939	hypothetical prote
21	295	7.9	738	2 T19938	hypothetical prote
22	290.5	7.8	1119	2 AD1822	leucine-rich repea
23	289.5	7.8	594	2 T23841	hypothetical prote
24	287.5	7.7	458	2 T19941	hypothetical prote
25	286.5	7.7	961	2 T23395	hypothetical prote
26	285.5	7.7	610	2 T23836	hypothetical prote
27	285.5	7.7	1385	2 T13887	tlr protein - fru1
28	283	7.6	1066	2 T15864	hypothetical prote
29	283	7.6	1389	2 T13852	gene wheeler prote

30	280.5	7.5	662	2 S42799	garp precursor - h
31	276.5	7.4	357	2 S24317	decorin precursor
32	274.5	7.4	575	2 T29972	hypothetical prote
33	273	7.3	1112	2 T10504	disease resistance
34	272	7.3	1134	1 A29944	chaoptin precursor
35	270.5	7.2	420	2 A53531	oncofetal trophobl
36	268	7.2	354	2 A53454	decorin precursor
37	268	7.2	360	2 S06280	decorin precursor
38	266	7.1	1097	2 A29943	Toll protein precu
39	266	7.1	1535	2 S46224	peroxidasin - fru1
40	263	7.0	1016	2 T30553	disease resistance
41	262.5	7.0	562	2 T34319	hypothetical prote
42	262.5	7.0	1051	2 T13174	gp150 protein - fr
43	261	7.0	359	1 NBH0C8	decorin precursor
44	261	7.0	360	2 I47020	decorin - rabbit
45	258	6.9	354	2 S29145	decorin precursor

ALIGNMENTS

RESULT 1
JC7763 neuronal leucine-rich repeat protein-3 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
A:Accession: JC7763
R:Fukunishi, K.; Matsuo, Y.; Kitahara, C.; Kuchino, Y.; Tsuda, H.
Biochem. Biophys. Res. Commun. 287, 257-263, 2001
A:Title: Rat neuronal leucine-rich repeat protein-3: Cloning and regulation of the ge
A:Reference number: JC7763; PMID:11549284
A:Contents: Fibrosarcoma cells
A:Accession: JC7763
A:Molecule type: mRNA
A:Residues: 1-707 <FWK>
A:Cross-references: GB:AF291437
C:Comment: This protein, a new member of the neuronal leucine-rich repeat protein fam
in protein-protein interaction and functions as a cell adhesion molecule or soluble I
C:Genetics:
A:Gene: nlrr-3
C:Keywords: cell adhesion

Query Match	44.2%	Score 1649.5	DB 2	Length 707
Best Local Similarity	47.5%	Pred. No. 4.5e-105		
Matches 337	Conservative 112	Mismatches 229	Indels 31	Gaps 7
OY	6	APLLAWAG--ATATPVVPMHVPCCPQACQIRPWYTPRSYREATVDCNDLFTAV	63	
DB	4	APLDIHVLGLDALTALVQAGDKKVDPCQICETIRPWTFRSIVWEASTVDCNDLGLNF	63	
OY	64	PPALPACTGTLTLQNSIVRVDSGLGYANTTELDLSONSFSDARDCDFALPQLSLH	123	
DB	64	PARLPACTGTLTLQNTNNINRIEHS-TDFPVNLTGDLSSNNLSVTNNIVKMSQLLSVY	122	
OY	124	LEENQLRLDHPSPAGLASHQELLYLNHNLVRIAPRAFSGLSNLRLNLNSLRAISR	183	
DB	123	LEENKLELEPKLYGSLNQLVYVHNLSSAISCAGVGLNLRLLNNSNRLOMNSK	182	
OY	184	WFEMLPMLLEIMTGNKVADILDMNFRPLANRSLVLAGMNLRETSYDLAEGLOSLSLS	243	
DB	183	WFEMLPMLLEIMTGNKVADILDMNFRPLANRSLVLAGMNLRETSYDLAEGLOSLSLS	242	
OY	244	FYDNLARVRPRALRQVPGKPLDKNPLNLRQVGDPAANNILKELGNNMELVSIIDK	303	
DB	243	FYDNLARVRPRALRQVPGKPLDKNPLNLRQVGDPAANNILKELGNNMELVSIIDK	302	
OY	304	FALVNEBELTKLDTNNRSLFTHPRAFHRLPOMETLMNNNAVALHQQTVESLPMIQE	363	
DB	303	LAVNLEPDLKRIATANNRSLYTHPRAFHRLPOMETLMNNNAVALHQQTVESLPMIQE	362	
OY	364	VGLHGNIRDCDVTIRANATGFRVRIEFGSLCAEPDGLQRLPVREVPFEMDHLPL	423	

Db 363 ISHSNPICDCVIRMINNKNTINIFEMPEDSLFCVDPPEFGQGNVRQVHRDMMEICLPL 422
 QY 424 ISPRSPISLOVASGESVYLHCRALEPEPEIYWTAGLIRPAHAGRRYRPGTLE 483
 Db 423 IAPSPSLIDYADSYSLHCRATAPPEIYWTISGKRLDPNTIREFYHSECTLD 482
 QY 484 LRRVTAEEAGLYTCAONLGVADITVSVVGRALLDPGRDEGGLELRYOETHPIHLL 543
 Db 483 IRGIFPEEGGLYTCTATNTLVGADKSIKIKYGFVPO---DNNGSLIKIRDIRANSVLY 539
 QY 544 SVWTPPNVTSTNLTVSSASLSRGATATLAPLPGTSHSYNTRTLQATEYMACLOVAFPD 603
 Db 540 SMKANIKILKSSVKYKTAFTKEDSOAASAKIPSDVKYTNLTKPSTETKICIDIPITY 599
 QY 604 AHTOLACVART-----KEATSCH---RALDRPGLIATLAVLLAAGIAAHLGT 652
 Db 600 OKRKQCCVWTTKSLHGGKENGKSHTEFVACVGLLGIIGVCL-----FGCVSEQGN 653
 QY 653 GQPRKQVGRPLPAPMAFWMGASPSVRSAPLVLPMNGRKLPRSE 701
 Db 654 CENEHSYTNHCHKPTLAF-----SELYPPLINIMESKKEFPASLE 694

RESULT 2
 glial cell membrane glycoprotein LIG-1 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change 05-Nov-1999
 C:Accession: A58532
 R:Suzuki, Y., Sato, N., Tohyama, M., Wanaka, A., Takagi, T.
 J. Biol. Chem. 271, 22322-22327, 1996
 A:Title: cDNA cloning of a novel membrane glycoprotein that is expressed specifically in A:Accession number: A58532; MUID:96394313
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1091 <SU2>
 A:Cross-references: GB:D78572; NID:g1545806; PIDN:BA11416.1; PID:g1545807
 C:Superfamily: Leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-ter
 F:36-61/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
 F:71-94/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
 F:95-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
 F:118-141/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
 F:142-165/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
 F:166-189/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
 F:191-213/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
 F:214-231/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
 F:238-261/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
 F:262-285/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>
 F:286-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>
 F:310-333/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>
 F:334-357/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>
 F:358-381/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>
 F:385-408/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>
 F:409-432/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR16>
 F:440-485/Domain: proteoglycan carboxyl-terminal homology <CCH>

Query Match 11.8%; Score 442; DB 2; Length 1091;
 Best Local Similarity 23.2%; Pred. No. 4.8e-22;
 Matches 169; Conservative 106; Mismatches 253; Indels 198; Gaps 22;

QY 5 VAPLLAMV-----AGATATVPVWHPVPCPOCAC-----01 37
 Db 14 LAPRLMLLMLLLQWPESAGAQARP-----RAFCACACGACAGSLDCSGRLATLPRDL 68
 QY 38 RMYWYPRS---SYREATIVD-----CNLFITAVPALPA--GTQITLLDS 78
 Db 69 PSM--TRSTNLSYNRLSSIDSAFEDITNQEVYLNSTNLTALPSLGTASIGVSLFLOH 126
 QY 79 NSIVRDOSELGYLANLELDLSONSFSDARDCFHALLPOLLSLHLEENOLFLEHSPFA 138
 Db 127 NKLISVDGSQLKSYLSLEVLDDLSSNNITEIRSSCFPMGLIRRELINLASRISTLES GARD 186

QY 139 GLA-SLOELYLNHNLQYRLAPRAFSGLSNLRLHLNSNLRAIDSRWEMLPNLEIMIG 197
 Db 187 GLSSSLTLRLSKNRRTQUPVAFK-LPRLIDLDNRKRIRILIGLTFGGIDSLSEVLRQ 245
 QY 198 GKNVDAITLDMNFRPLNLRSLVLAGMNLREISDVALBGLQLESLSFYDQALVRPR-- 254
 Db 246 RNNISRLTDGAFWGLSKMHVHLLEVNSLYEVNSGLYGLTALHOLHLSNNSISRIHQDW 305
 QY 255 -----RALDQVGLKFLDKNKNPLQORVPGDPANMLHKEELGN 293
 Db 306 SFQCKLHELILSNLNLTRIDESLAEISLSTRSHAIHIEGAFKGLKSLRYVLDL 365
 QY 294 NMEELYSID--KEALVNPDELTKDITNNPRLSFTHPRAFHILPOMETMLNNAALSALH 351
 Db 366 HNEISGTEIDTSGAFGLNLKSLTLFGN-KIKSVAKRAFGLESLHLNGENAIRSYQ 424
 QY 352 QQTVESLPNLQEVGLHGNPRLDCVYRW-----ANNT-----GTRVRFI 390
 Db 425 FDAFAKKNLKELYISESEFLDCDQKMLPRLMGRLQATVYATCAHPESLKGQSTFSV 484
 QY 391 EPOSTLCAEPP-----DL 403
 Db 485 LPDSFVCDPEPKPQIITQPEETMAVVGKDIRFTCSAASSSSPMTFAKKDNEVLAMDM 544
 QY 404 OR-----LPVREVPF-----REMTDHCPLIS 425
 Db 545 ENRAHRAQDGEVMEYTTLHLRHTFGHEGRYOCITNHRGSTRYSHKRLTVNLP--S 602
 QY 426 PRSPPSLOVASGESVYLHCRALEPEPEIYWTAGLIRPAHAGRRYRYP-EGTLEL 484
 Db 603 FTKIPHDIAIRCTVTRLECAATGHPNPQIAMQXGKIDF--PAARERMYHMPDDVFFI 661
 QY 485 RRYTAEEAGLYTCAONLGV--ADRTKTSVVVGRALLDPGRDE-----GQGLLELYOETHP 538
 Db 662 TDVKIDMGVYSTQASNSAGSVSANATLFTVETPSLANPLDPRVYVGYEYARQCRATGS 721

RESULT 3
 A41915
 Insulin-like growth factor-binding complex acid-labile chain precursor - human
 N:Alternate names: Acid-labile Subunit (ALS)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
 C:Accession: A41915
 R:Leong, S.R.; Baxter, R.C.; Camerato, T.; Dai, J.; Wood, W.I.
 Mol. Endocrinol. 6, 870-876, 1992
 A:Title: Structure and functional expression of the acid-labile subunit of the insulin A:Reference number: A41915; MUID:92357025
 A:Accession: A41915
 A:Status: preliminary
 A:Molecule type: mRNA; protein
 A:Residues: 1-605 <LEO>
 A:Cross-references: GB:M86826; NID:g184807; PIDN:AAA36047.1; PID:g184808
 A:Experimental source: liver
 A:Note: sequence extracted from NCBI database (NCBIPI:110171)
 C:Superfamily: Leucine-rich alpha-2-glycoprotein repeat homology
 F:75-99/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
 F:99-122/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
 F:123-146/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
 F:147-170/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
 F:171-194/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
 F:195-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
 F:219-242/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
 F:243-266/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
 F:267-290/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
 F:291-314/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>
 F:315-338/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>
 F:339-362/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>

us-09-905-088a-245.rpr

```

Db      465  RCASRRRLANKRIGQIKSKKFRCSAKQEQFTLP--GTEDYHUNSECTSDVACPCHKRC--- 519
Oy      451  PEPEIYWTYPAGLRRLT--PAHAGRRRYRYPGTELELR 485
Db      520  ---EASVYECGSLKLSKIPER-----IPQSTTELR 546

RESULT      9
JEO176
orphan G protein-coupled receptor precursor - human
C;Species: Homo sapiens (man)
C;Date: 03-Jul-1998 #sequence.revision 10-Jul-1998 #text.change 21-Jul-2000

```

B:Recommendation: Biophys. Res. Commun. 247, 266-270, 1998
 A:Title: Identification and cloning of an orphan G protein-coupled receptor of the G12 family
 A:Reference number: JF0176; MUID:98308104
 A:Accession: JF0176
 A:Molecule type: mRNA
 A:Residues: 1-907 <MCD>
 A:Cross-references: GB:AF062006; NID:g3366801; PID:g3366802
 C:Comment: This protein is a receptor for a novel class of glycoprotein ligands.
 C:Genetics:
 A:Gene: HG38
 A:Map position: 12q22-23
 F:1-71/Domain: signal sequence #status predicted <SIG>
 F:562-583/Domain: transmembrane #status predicted <TM1>
 F:594-616/Domain: transmembrane #status predicted <TM2>
 F:639-660/Domain: transmembrane #status predicted <TM3>
 F:681-701/Domain: transmembrane #status predicted <TM4>
 F:725-744/Domain: transmembrane #status predicted <TM5>
 F:768-791/Domain: transmembrane #status predicted <TM6>
 F:803-824/Domain: transmembrane #status predicted <TM7>

Query Match	9.6%;	Score 360;	DB 2;	Length 907;
Best Local Similarity	27.4%;	Pred. No. 1.5e-16;		
Matches 134;	Conservative 70;	Mismatches 189;	Indels 96;	Gaps 14

[illegible]


```

QY 209 -----FRLAHLRLSYLAGMNLREISDYALGLOSLESISTYDNOIARY----- 252
Db 301 AITSLRPRTFKDLHLEELQGLHNRIRQLGKTEKFFGLQLLEVLTLTNDNOIHKKVKAFFG 360
QY 253 -----PRALEQVPGLKFLDLNKNPQLQRYVPGCFANMLHKEKGLNM 295
Db 361 LFNVAVMNLGSGCLSLPRHVFQGLGRHLSLHESCLGRITLHFAGLSGRLRFLNR- 419
QY 296 EELVYSIDKFALYNLEPLKQLDITNNPRLSFTHPRAFHILPQMETLMLNNALSALHQ--- 352
Db 420 NSISIEEQSLAGLELLELDLTAN-OLTHLPRLFQGLGLELYLLSNNOQLTMSDEVL 478
QY 353 -----OTVESLPNLOEVLGKNPFR 372
Db 479 GPLQAFWDLDSHNLLETPAEGLFSSLGRLRYLNRNNSLGTFVFPQGLERLMDLANPMD 538
QY 373 CDCVTR-----WANKTGRVRFIEP-----OSTCAEPPDIQLRPVREVPFR 414
Db 539 CSCPRKALNDELQNPQVVPREFVQVCEGDDCQPYTYNNITCAGPANVSGDLRDI--- 595
QY 415 EMTDCHLPLISPRSPPSIQVAGSGEMVLC 445
Db 596 -----SETLRFVHC 603

RESULT 13
B3665
silt protein 2 precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 19-May-2000
C:Accession: B3665
R:Rothenberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A:Title: silt: an extracellular protein necessary for development of midline glia and co
A:Reference number: A3665; MUID:91099665

```

[illegible]

slit protein precursor - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 17-Nov-2000
C:Accession: A36665; A31640; S13523
R:Rothenberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A>Title: slit: an extracellular protein necessary for development of midline glia and cell
A:Reference number: A36665; MUID:91099665
A:Accession: A36665
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1480 <RO1>
A:Cross-references: GB:535959; NID:98614; PIDN:CAA37910.1; PID:98615
R:Rothenberg, J.M.; Hattley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
Cell 55, 1047-1059, 1988
A>Title: slit: An EGF-homologous locus of *D. melanogaster* involved in the development of
A:Reference number: A31640; MUID:89077533
A:Accession: A31640
A:Molecule type: DNA
A:Residues: 881-1182, 'G', 1185-1404, 'GT', 1463-1464, 'YHA', <RO2>
A:Cross-references: GB:M23543; NID:9340939; PID:9514357
C:Genetics:
A:Gene: FlyBase:slit
A:Cross-references: FlyBase:FBgn0003425
A:Introns: 1351/3
C:Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glyc
C:Keywords: alternative splicing; growth factor
F:66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F:101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F:125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F:149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F:173-186/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F:197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F:228-272/Domain: proteoglycan carboxyl-terminal homology <PCST1>
F:288-313/Domain: proteoglycan amino-terminal homology <PAH2>
F:323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F:347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F:371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F:395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F:419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>
F:450-494/Domain: proteoglycan carboxyl-terminal homology <PCST2>
F:512-537/Domain: proteoglycan amino-terminal homology <PAH3>
F:547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>
F:572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>
F:596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>
F:620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>
F:651-699/Domain: proteoglycan carboxyl-terminal homology <PCST3>
F:708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F:743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>
F:767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR16>
F:791-814/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR17>
F:815-838/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR18>
F:846-890/Domain: proteoglycan carboxyl-terminal homology <PCST4>
F:1028-1061/Domain: EGF homology <EGF2>
F:1068-1099/Domain: EGF homology <EGF2>
F:1115-1148/Domain: EGF homology <EGF1>

QY 209 FRPLAN-LRSLVLAGANLR-----EISDYALBELQSLF----- 240

Db 661 LAWEAEVCRKRKSLNGGAAFGAPSKVADVQIKPLPSEKCSSENSGCLDGCTPSPCT 720

QY 241 ----SLSEFYNOIARVFRRALEQVPG-LKFLDANKKPLQVRVGGPDANMLHEKELGLANM 295

Db 721 CTGVVACSNRQQLKEIFR-----GIPIAETSELYLESENDEIDQIHYERIRHRSILTRDLISN- 775

QY 296 EEIVSIDKFAVLVNPETLKLDITNNPRLSFIHFRATFHHLPQMETLMLNNALSAHQGV 355

Db 776 NOITLISNYFANITLTKLSTLIISYN-KLQCLQRLHSLGNNLRVSLHGNRISMPEGSEF 834

QY 356 EESLNIQLEVGILHGPCRDCCVIRANATGTGRVFIEPOSTILCAEPDLQRLPYREVP 412

Db 835 EDLKSILTHIALGSLNPFLYCDGGLKWF-SDWIKIDYVEGRICARCAEPDMMDKILSTP 890

RESULT 15

A60164

platelet membrane glycoprotein V precursor - human

C:Species: Homo sapiens (man)

C:Date: 12-Jan-1993 #Sequence-Revision 24-Feb-1994 #text-change 05-Nov-1999

C:Accession: A48030; A60164; A35483; B35483; C35483; A60432; A47507; S34329

R:Rizana, F.; Morales, M.; de La Sallie, C.; Cazenave, J.P.; Clementson, K.J.; Shimomura

J. Biol. Chem. 268, 20801-20807, 1993

A:Title: Cloning and characterization of the gene encoding the human platelet glycopr

A:Reference number: A48030; MUID:94012616

A:Accession: A48030

A:Molecule type: DNA

A:Residues: 1-560 <L2>

A:Cross-references: EMBL:Z23091; NID:g312501; PIDN:CAA80637.1; PID:g312502

R:Shimomura, T.; Fujimura, K.; Maehama, S.; Takemoto, M.; Oda, K.; Fujimoto, T.; Oyam

Blood 75, 2349-2356, 1990

A:Title: Rapid purification and characterization of human platelet glycoprotein V: th

A:Reference number: A60164; MUID:90275263

A:Accession: A60164

A:Molecule type: protein

A:Residues: 365-384, 'X', 386-390, 'X', 392-395, 'X', 397, 188-208, 'T', 210, 27-50, 'X', 52-53, 1

, 'XX', 108, 'T', 61-72, 'TK', 75-77, 'V', 56-57, 'G', 479-487, 'X', 489-498, 'X', 500, 'X', 502-503,

R:Roth, G.J.; Church, T.A.; McWilliams, B.A.; Williams, S.A.

Biochem. Biophys. Res. Commun. 170, 153-161, 1990

A:Title: Human platelet glycoprotein V: a surface leucine-rich glycoprotein related t

A:Reference number: A35483; MUID:90321220

A:Accession: A35483

A:Molecule type: protein

A:Residues: 145-166, 'T', 168-169, 'X', 171-172 <RO>

A>Note: this proteolytic fragment was designated peptide M392

A:Accession: B35483

A:Molecule type: protein

A:Residues: 121-129, 'W', 131-135, 466-468, 'X', 470 <RO2>

A>Note: this material was designated peptide M393 but may contain two peptides

A:Accession: C35483

A:Molecule type: protein

A:Residues: 252-266, 'H', 268-272, 'X', 274-279, 'T', 281-284, 'T', 286 <RO3>

A>Note: this proteolytic fragment was designated peptide M401

R:Zafar, R.S.; Walz, D.A.

Thromb. Res. 53, 31-44, 1989

A:Title: Platelet membrane glycoprotein V: characterization of the thrombin-sensitive

A:Reference number: A60432; MUID:89162331

A:Accession: A60432

A:Molecule type: protein

A:Residues: 477-478, 'FX', 481-485, 'E', 487, 'V', 489-492, 'NQ', 495, 'E', 497-498 <ZAF>

R:Hickey, M.J.; Hagen, F.S.; Yag1, M.; Risch, J.

Proc. Natl. Acad. Sci. U.S.A. 90, 8327-8331, 1993

A:Title: Human platelet glycoprotein V: characterization of the polypeptide and the r

A:Reference number: A47507; MUID:93319348

A:Accession: A47507

A:Status: preliminary; translated from GB/EMBL/DBU

A:Molecule type: mRNA

A:Residues: 1-560 <RES>

A:Cross-references: GB:I11238; NID:g388759; PIDN:AA03069.1; PID:g388760

C:Comment: This platelet membrane protein is a substrate for thrombin.

C:Comment: The amino end of the intact protein is blocked.
C:Comment: This protein is absent in Bernard-Soulier syndrome.
C:Genetics:
A:Gene: GDB:GP5
A:Cross-references: GDB:230236; OMIM:173511
A:Map position: 5pter-5qter
C:Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
C:Keywords: blocked amino end; glycoprotein; platelet; tandem repeat; transmembrane prot

Query Match 8.6%; Score 321; DB 2; Length 560;
Best Local Similarity 26.6%; Pred. No. 3,5e-14;
Matches 141; Conservative 57; Mismatches 227; Indels 106; Gaps 16;

```

OY 9 LLAVVAGATATVPVPMHVPQPCQACQIRPWYTPRSSYREATTVCNDLF-LTAVPPAL 67
DB 6 LICAVALGILRAQP-----FPCPPACKC-----VERDAQCSSGDVARISAL--GI 48
OY 68 PAG--TOTLLQSNSTIVRVDSGLYLANLTPELDLSONSFSDARCDPFAHPOLSLHLE 126
DB 49 PTNLTHILFEGMGRGVLSQSSFSG-MTVLQRLMISDSHISAVAPGTFSDLIKIKTLRLSR 107
OY 127 NQLRLEDHSPAGLASIQELYLNHNQLYRIAPRAFSGLNLRHLNLSNLLRAIDSRWE 186
DB 108 NKITHLPQALDKWVLEQLFDHNALRGIDQNMFKLVNQELALNQNLDFLPASLFT 167
OY 187 MLPNLEILMIGSNKVDAIIDMNFRLANLSIVYLAGMNLREISDYALEGLQSLSESFYD 246
DB 168 NLENKLLDLSGNNLTHTLPKGLGAQAKLERLHLNRLVSLDSGLNSIGALTLELQFHR 227
OY 247 NQARVPRRALEQVPGIKFLDLNK-----NPLQRVGPGDPA 282
DB 228 NHRSIAGAFDRLPNLSLTLNHLAFLPSALFLHSHNLTLLTFENPLAELPGVLF 287
OY 283 NMLHKEGLNMMELVSIKFAVNPPELTKDITNNPRLSFIHPRAFHHPQMETML 342
DB 288 EMGGLQELMLNR-TQLRTLPAAAFRLNRLRYLGLVTLSPRLSALPGAFQGLGELQYLA 346
OY 343 NNNALSAL-----HOQ-----T 354
DB 347 HSNGLTALPDGLRGGLKRGVSRNRRLALPRALFRNLSSLESYQLDHNQLETLPGDV 406
OY 355 VESLPNLOEVGLHGNPIRDCD---VIRMANATGTRVRETEPOSTLCAEPPDQLRLVRE 410
DB 407 FGALPRLTEVLIGHNSWRCDCGLCPFLGWLRQHLGLVGCGEPPR--CAGPGAHAGLPWA 464
OY 411 VPFRMTDHCPLISPRSFPSLQVAGSESMVLCRALAEPEPEY-WVTP 460
DB 465 LPGGDA--EC---PGPRGPPPPRAADSSSEAPVH-PALAPNSSEPPWVAQP 509

```

Search completed: August 31, 2002, 14:36:40
Job time: 102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:35:28 ; Search time 13.6 Seconds
(without alignments)
2029.929 Million cell updates/sec

Title: US-09-905-088a-245
Perfect score: 3732
Sequence: 1 MRLVAPLLAWAGATATV.....RKUPRSEGETLPPPLSQNS 713

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3718	99.6	713	GAC1_HUMAN	O75325 homo sapien
2	407.5	10.9	605	AUS_HUMAN	P35858 homo sapien
3	397.5	10.7	605	AUS_PAPHA	O02833 papio hamad
4	344	9.2	603	AUS_RAT	P35859 rattus norv
5	342.5	9.2	603	AUS_MOUSE	P70389 mus musculu
6	340.5	9.1	567	GPV_MOUSE	O06742 mus musculu
7	324	8.7	1480	SLIT_DROME	P24014 drosophila
8	321	8.6	560	GPV_HUMAN	P40197 homo sapien
9	314	8.4	905	TRR3_MOUSE	O95MD1 mus musculu
10	313.5	8.4	567	GPV_RAT	O08770 rattus norv
11	298.5	8.0	361	CHAD_BOVIN	O27972 bos taurus
12	287	8.0	536	CBP8_HUMAN	P22792 homo sapien
13	287	7.7	904	TRR3_HUMAN	O15455 homo sapien
14	280.5	7.5	662	GARP_HUMAN	O14392 homo sapien
15	279	7.5	782	CHAO_TRICA	P83965 trilbolium c
16	276.5	7.4	357	PGS2_CHICK	P28673 gallus galli
17	272	7.3	1315	CHAO_DROME	P12024 drosophila
18	270	7.2	360	PGS2_PIG	O95MD1 mus musculu
19	270	7.2	1050	TRR7_MOUSE	P58681 mus musculu
20	269.5	7.2	1032	TRR9_MOUSE	O95MD1 mus musculu
21	268.5	7.2	1032	TRR9_HUMAN	O95MD1 mus musculu
22	268	7.2	354	PGS2_MOUSE	O95MD1 mus musculu
23	268	7.2	360	PGS2_BOVIN	P28658 mus musculu
24	266	7.1	1097	TOLL_DROME	P21793 bos taurus
25	261	7.0	359	PGS2_DROME	P08953 drosophila
26	261	7.0	360	PGS2_HUMAN	P08953 drosophila
27	260	7.0	977	Y848_HUMAN	O28888 cyrtocolagus
28	259	6.9	331	PLIB_AGRBL	O94933 homo sapien
29	258	6.9	354	PGS2_RAT	O93223 agkistrodon
30	253	6.8	1049	TRR7_HUMAN	O01119 rattus norv
31	253	6.8	360	PGS2_CANFA	O95MD1 mus musculu
32	253	6.8	372	PGS1_HORSE	O29333 canis fami
33	252	6.8	1032	TRR8_MOUSE	P58682 mus musculu

ALIGNMENTS

```

RESULT 1
GAC1_HUMAN
AC 075325; STANDARD: PRT: 713 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glioma amplified on chromosome 1 protein precursor.
GN GAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Glial tumor;
RX MEDLINE=96324709; PubMed=9662332;
RA Malfroy B., Almeida A., Zhu X.X., Vogt N., Tyagi R., Mulleris M.,
RA Dutrillaux A.-M., Dutrillaux B., Ross D., Hanash S.;
RT "GAC1, a new member of the leucine-rich repeat superfamily on
RT chromosome band 1q32.1, is amplified and overexpressed in malignant
RT gliomas."
RL Oncogene 16:2997-3002(1998).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: OVERAMPLIFIED IN MALIGNANT GLIOMAS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF030435; AAC39792.1; -.
DR MIM; 605492; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Nterm.
DR InterPro; IPR000372; LRR_Cterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_tyr.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00560; LRR_10.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00408; ICG2; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 2.

```

KM Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat;
 KW Leucine-rich repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 713 GLIOMA AMPLIFIED ON CHROMOSOME 1 PROTEIN.
 FT DOMAIN 19 630 EXTRACELLULAR (POTENTIAL).
 FT TRNSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 713 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 92 115 LRR 1.
 FT REPEAT 116 139 LRR 2.
 FT REPEAT 140 163 LRR 3.
 FT REPEAT 165 187 LRR 4.
 FT REPEAT 188 211 LRR 5.
 FT REPEAT 213 235 LRR 6.
 FT REPEAT 236 259 LRR 7.
 FT REPEAT 261 283 LRR 8.
 FT REPEAT 309 333 LRR 9.
 FT REPEAT 334 357 LRR 10.
 FT REPEAT 359 385 LRR 11.
 FT REPEAT 438 504 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 445 497 BY SIMILARITY.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 555 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 713 AA; 78798 MW; EC8BC0FD240C9396 CRC64;

Query Match 99.6%; Score 3718; DB 1; Length 713;
 Best Local Similarity 99.7%; Pred. No. 2,1e-261;
 Matches 711; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MRLVAVPILLAWAGATVAVPVVPMVHPCPCACQIRPMVPRSSSYREATTVCNDLFL 60
 DB 1 MRLVAVPILLAWAGATVAVPVVPMVHPCPCACQIRPMVPRSSSYREATTVCNDLFL 60
 OY 61 TAVPPALPAGTQTLILQSSNSIVRVDOSLGYLANITLTDLSQNSFSDARCDFFALPOLL 120
 DB 61 TAVPPALPAGTQTLILQSSNSIVRVDOSLGYLANITLTDLSQNSFSDARCDFFALPOLL 120
 OY 121 SHALENOITRLLEDHSEFAGLASIQELYLNHNDLYRAPAFSGSLNLRHLNSMLTRAI 180
 DB 121 SHALENOITRLLEDHSEFAGLASIQELYLNHNDLYRAPAFSGSLNLRHLNSMLTRAI 180
 OY 181 DSRFEMFLPNLEITLMIGKKNVDAILDMNFRPLANLRSVLACMNLREISDYALEGQSLE 240
 DB 181 DSRFEMFLPNLEITLMIGKKNVDAILDMNFRPLANLRSVLACMNLREISDYALEGQSLE 240
 OY 241 SLSFYDNOGLARVPRRALEQVPGIKFLDLNKNPLQRYGPGDFPANNMLHLKELGINNMEELVS 300
 DB 241 SLSFYDNOGLARVPRRALEQVPGIKFLDLNKNPLQRYGPGDFPANNMLHLKELGINNMEELVS 300
 OY 301 IDEFALVNPBELTKLDITNNPRLSFTHRAPFHHLPOMETLMNNNALSLAHQOYESLPN 360
 DB 301 IDEFALVNPBELTKLDITNNPRLSFTHRAPFHHLPOMETLMNNNALSLAHQOYESLPN 360
 OY 361 IQEVGLHGNPIRCDCVIRMANATGTRVRELEPOSTLCAEPDQLRVREVPREKTDHC 420
 DB 361 IQEVGLHGNPIRCDCVIRMANATGTRVRELEPOSTLCAEPDQLRVREVPREKTDHC 420
 OY 421 LPLISRSRPPSLQVASGESMTVHCRALAEPRPEIYWTTPAGRLCPRAAGRRYRYRPG 480
 DB 421 LPLISRSRPPSLQVASGESMTVHCRALAEPRPEIYWTTPAGRLCPRAAGRRYRYRPG 480
 OY 481 TLELRRTAEAGLYTCVANOILGADTKTVYVVGRRALQOPGRDESGLELRYOETHRNH 540
 DB 481 TLELRRTAEAGLYTCVANOILGADTKTVYVVGRRALQOPGRDESGLELRYOETHRNH 540
 OY 541 ILLSWTTPPTVSTNLTWSSASSLRQOGATALARLPRGTHSYNITRLQATEYMACLOYA 600
 DB 541 ILLSWTTPPTVSTNLTWSSASSLRQOGATALARLPRGTHSYNITRLQATEYMACLOYA 600
 OY 601 FADAHTQLACVMARTKEATSCRRALDRPGLIALLAVILLAAAGLAHLGTGQPRKGVG 660
 DB 601 FADAHTQLACVMARTKEATSCRRALDRPGLIALLAVILLAAAGLAHLGTGQPRKGVG 660

DB 601 FADAHTQLACVMARTKEATSCRRALDRPGLIALLAVILLAAAGLAHLGTGQPRKGVG 660
 OY 661 GRRLPPANAFWGSAPSVRYVSAPLVLPNPNGRKLPSSSEGETLLPLPSQNS 713
 DB 661 GRRLPPANAFWGSAPSVRYVSAPLVLPNPNGRKLPSSSEGETLLPLPSQNS 713
 RESULT 2
 ALS_HUMAN STANDARD; PRT; 605 AA.
 ID ALS_HUMAN
 AC P35858;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Insulin-like growth factor binding protein complex acid labile chain precursor (ALS).
 GN IGFBP3 OR ALS.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=92357025; PubMed=1379671;
 RA Leong S.R., Baxter R.C., Camerato T., Dai J., Wood W.I.;
 RT "Structure and functional expression of the acid-labile subunit of the insulin-like growth factor-binding protein complex.";
 RL Mol. Endocrinol. 6:870-876(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Frankland J.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP MEDLINE=89308584; PubMed=2473065;
 RA Baxter R.C., Martin J.L., Beniac V.A.;
 RT "High molecular weight insulin-like growth factor binding protein complex. Purification and properties of the acid-labile subunit from human serum.";
 RL J. Biol. Chem. 264:11843-11848(1989).
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT IN PROTEIN COMPLEXES. RECEPTOR-LIGAND BINDING OR CELL ADHESION.
 CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH IGF-1 OR IGF-II AND IGFBP-3.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M86826; AAA36047.1; .
 CC EMBL; AL031724; CAC36078.1; .
 CC PIR; A41915; A41915.
 CC HSSP; P23945; LXUN.
 CC MIM; 601489; .
 CC InterPro; IPR001611; LRR.
 CC InterPro; IPR000483; LRR_Cterm.
 CC InterPro; IPR000372; LRR_Nterm.
 CC InterPro; IPR003592; LRR_out.
 CC InterPro; IPR003591; LRR_tyr.
 CC Pfam; PF00560; LRR_19.
 CC Pfam; PF01463; LRRCT; 1.
 CC Pfam; PF01462; LRRNT; 1.
 CC PRINTS; PR00019; LEURICHRPT.
 CC SMART; SM00370; LRR; 2.
 CC SMART; SM00882; LRRCT; 1.

Query Match	10.9%;	Score 407.5;	DB 1;	Length 605;
Best Local Similarity	25.2%;	Pred. No. 6.5e-22;		
Matches 152;	Conservative 72;	Mismatches 173;	Indels 205;	Gaps 16;

```

QY      3 LVAAPLLAMWA-----GATIAVPVPMWHPDPOCACQIRPMWTPPRSY--REAWT 52
           | : ||| ||||
Db       8 LAIALILLSWVALIGRSLLEGADPGTPEABGAPACPAAYC-----SYDDADELS 57
           | : ||| ||||
QY      53 VDCNDLEFLAVPALPAGTQTQLLOSNISIVR-----DOSELGY-          91
           | : ||| ||||
Db       58 VFCCSRNLTFRPDVGPGGTQALMIDGNNTSSVPPAFQNLSSLGLFNLNAGOGOLSEPOA 117
           | : ||| ||||
QY      92 ----LANTFLDISONSFECDARCDFFHALPOLLSLHEENQTRLE-----    133
           | : ||| ||||
Db      118 LGLENLCHLHNERNQNSLALGTFTATPALASLGTSNNRLSREDGLFEGLSLMDLNTL 177
           | : ||| ||||
QY      134 -----DHSPALIASIOELYLNHQLRIRAPRAFSGLSMLTLHLINSULRAIDSRW 184
           | : ||| ||||
Db      178 GMSNSTAVLPDAAFKGLGSRLRYLAGNRKIAYIOPALPFESGLAEURELDSRRALARAIKANV 237
           | : ||| ||||
QY      185 FEMLPNELETIMIGNKYDAI-----LDNN-----                208
           | : ||| ||||
Db      238 FVOLPRLOKTYLDRMLAAVAPGAFLGKALKRMVLDSHNHRVAGLEDDEFPFGILGRVRL 297
           | : ||| ||||
QY      209 -----FRPLANLSVLYAGNNLRREISDYALEBGLQSLESLSFYDNOLARV--- 352
           | : ||| ||||
Db      298 SHNAIASLRPTFKDLHFLELOIGHNRIROLAERSESFEGLGLEVTYLDHNQJOLEVRAGA 357
           | : ||| ||||
QY      253 -----PRALBEVDELKFIDLKNKPRLORVGPBGFAMMLHKELGL 292
           | : ||| ||||
Db      358 FLGLTNVAVMNLSGNCRLRNIPEDVYFPGKLKLSHLIEGSCGJRIRPHFTFGLSGLRFL 417
           | : ||| ||||
QY      293 NMMEELVSIDKFAVLNIPETLKLDITNNRPSFIHRPFAFHNPOMETLMNNNALSAI-- 350
           | : ||| ||||
Db      418 KD-NGIWVGEOSIMGLAEILELDTLSN-QLTHLPRLRFEGGKGXEYLLSRNLALBPRA 475
           | : ||| ||||
QY      351 -----HOQTVESLPN-----LOEVGLHG 368
           | : ||| ||||

```

```

Db      476 DALGPIQRAWMDVSHNRLEALPNSLLAPGRRLRYLSLRNLSLRTTTPDPGLERLMLEG 535
Qy      369 NPIRRCDCVIR---WANANGTRV-RTIEP-----QSTFLCAEPDLORLPYR 409
Db      536 NPMDCCGPIKLARDPFLONPSAVPRFVOAICGSDGDCQDPAYTYNNITCASPEVGLDLR 595
Qy      410 EV 411
Db      596 DL 597

RESULT 3
ALS_PAPHA ID ALS_PAPHA STANDARD: PRT: 605 AA.
AC 002833:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
DE precursor (ALS).
GN IGFALS OR ALS.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxId=9557;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97040714; PubMed=8886027;
RA Delhanty P., Baxter R.C.;
RT "The cloning and expression of the baboon acid-labile subunit of the
RT insulin-like growth factor binding protein complex.";
RL Biochem. Biophys. Res. Commun. 227:897-902(1996).
CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH
CC IGF-1 OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S83462; -; NOT_ANNOTATED_CDS.
DR HSSP: P23945; 1XUN.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR Pfam: PF00560; LRR_19.
DR Pfam: PF01463; LRRCT_1.
DR Pfam: PF01462; LRRNT_1.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00370; LRR_2.
DR SMART: SM00082; LRRCT_1.
DR SMART: SM00013; LRRNT_1.
DR SMART: SM00369; LRR_TYP_11.
KW Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 605
FT REPEAT 53 73
FT REPEAT 74 96
FT REPEAT 98 120
FT REPEAT 121 144
FT REPEAT 146 168
FT LRR 1
FT LRR 2
FT LRR 3
FT LRR 4
FT LRR 5
FT PROTEIN COMPLEX ACID LABILE CHAIN.

```

```

FT REPEAT 169 192 LRR 6.
FT REPEAT 194 216 LRR 7.
FT REPEAT 217 240 LRR 8.
FT REPEAT 242 264 LRR 9.
FT REPEAT 265 288 LRR 10.
FT REPEAT 289 312 LRR 11.
FT REPEAT 313 336 LRR 12.
FT REPEAT 338 360 LRR 13.
FT REPEAT 361 384 LRR 14.
FT REPEAT 386 408 LRR 15.
FT REPEAT 409 432 LRR 16.
FT REPEAT 433 456 LRR 17.
FT REPEAT 458 480 LRR 18.
FT REPEAT 482 504 LRR 19.
FT REPEAT 505 530 LRR 20.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 605 AA; 66110 MW; 9D71994625F23652 CRC64;

```

Query Match 10.7%; Score 397.5; DB 1; Length 605;
 Best Local Similarity 24.9%; Pred. No. 3,4e-21;
 Matches 150; Conservative 73; Mismatches 173; Indels 207; Gaps 16;

```

OY 3 LVAPELLAMVA-----GATATPVVPMVHPGPGACQIRPVYTRSSYREAT---T 52
DB 8 LALALLLSWALGPRSLGEGAEPTGEGAEPCATAC-----STDVENLS 57
OY 53 VDCNDLFTAVAPALPAGTOTLLQSNSTIVR-----DQS 87
DB 58 VCSSSNLRLRDPGIGQALMDSNNSISIPAPAFNLSSIAFLNLGGQGLSGEPQA 117
OY 88 ELGYLANTELDELSONSDARDCCPHALPOLLSHLEBNQTRLE----- 133
DB 118 LIG-LENLCHLHERQULSLAVGTFAYPALALGLSNRRSLRLEDGLFEGGLMMDLN 176
OY 134 -----DHSFAGLASLOELYLNHNOYLRIAPRAFSGLSNRLRLHNSLRAIDSR 183
DB 177 LGWNSLAVLPDAAFRGGLRELVLAGNRLAYLQPALSSGLAELELDLSRNLPAIKAN 236
OY 184 WFEMLPNEILMIGNKVDAL-----LDMN----- 208
DB 237 VEAQLPRIQKIVLDRNLIAAVAPAGFLGLKALRWLDLSHNRVAGLLEDTFPGGLGRVLR 296
OY 209 -----FRPLANLRSLIVAGANLREISDYALBGLQSLSSLSFYDQARV--- 252
DB 297 LSHNAIASLRPTFEDLHFLLEELQGHNRIRQLAERSFEGQLQLEVLTLDHNOQLEVKVG 356
OY 253 -----PRBALEOVPGLKFDLKNPNRQVPGVGFANMLHLKELG 291
DB 357 AFLGLTNVAVMNSGNCRLNPEQVFRGIGKLSHLBESSCGIRPHFAGLSGRRLF 416
OY 292 LNMEEVLSDKFAVNLDELTKLDTNNPRLSFTHPRAFNHLPMQETMLNNNALSAL- 350
DB 417 LKN-NGLVGIEEOSTMGLELLELDLTSN-QLTNHPHQFGIGKLELYLLSHNRLAELD 474
OY 351 -----HQQVIESL-----NILEVGGLH 367
DB 475 ADAALGLQARFMDVSHNRLEALPGSLASLGRLYRLNRNNSLRTFTPOPGLELWLE 534
OY 368 GNPTRCDQVTR---WANATGTRV-RFIEP-----OSTLCAEPPDQRLPV 408
DB 535 GANWDCSCPALKALRDPALQNPASVAPRFVQALICGDDCQPPVYTYNNITGASPEVAGDLD 594
OY 409 REV 411
DB 595 RDL 597

```

```

RESULT 4
ID ALS_RAT STANDARD: PRT; 603 AA.
AC P35859;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
DE precursor (ALS).
GN IGFALS OR ALS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93038676; PubMed=1384485;
RA Dai J., Baxter R.C.;
RT "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT growth factor binding protein complex.";
RL Biochem. Biophys. Res. Commun. 188:304-309(1992).
RN [2]
RP SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC STRAIN=MISTAR; Tissue=Serum;
RX MEDLINE=94130835; PubMed=7507839;
RA Baxter R.C., Dai J.;
RT "Purification and characterization of the acid-labile subunit of rat
RT serum insulin-like growth factor binding protein complex.";
RL Endocrinology 134:848-852(1994).
CC -1- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC CIRCULATING IGFs TO THE TISSUES.
CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH
CC IGF-1 OR IGF-II AND IGFBP-3.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC AND LIVER.
CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S46785; AAB23770.2; -.
CC PIR; JC1282; JC1282.
CC HSP; p23945; 1XUN.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEORICHPT.
DR SMART; SM00370; LRR; 5.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 9.
KW Glycoprotein; Leucine-rich repeat; Repeat; signal.
FT SIGNAL 1 23
FT CHAIN 24 603
FT REPEAT 52 73 LRR 1.
FT REPEAT 74 96 LRR 2.
FT REPEAT 98 120 LRR 3.
FT REPEAT 121 144 LRR 4.
FT REPEAT 146 168 LRR 5.
FT REPEAT 169 192 LRR 6.

```

```

FT REPEAT 194 216 LRR 7.
FT REPEAT 217 240 LRR 8.
FT REPEAT 242 264 LRR 9.
FT REPEAT 266 288 LRR 10.
FT REPEAT 289 312 LRR 11.
FT REPEAT 313 336 LRR 12.
FT REPEAT 337 360 LRR 13.
FT REPEAT 361 384 LRR 14.
FT REPEAT 386 408 LRR 15.
FT REPEAT 409 432 LRR 16.
FT REPEAT 433 456 LRR 17.
FT REPEAT 458 479 LRR 18.
FT REPEAT 479 504 LRR 19.
FT REPEAT 506 529 LRR 20.
FT REPEAT 543 566 LRR 21.
FT CARBOHYD 64 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 603 AA; 66811 MW; DCD7637D94A5037C CRC64;

```

```

Query Match 9.2%; Score 344; DB 1; Length 603;
Best Local Similarity 23.3%; Pred. No. 2,6e-17;
Matches 141; Conservative 63; Mismatches 194; Indels 208; Gaps 16;

```

```

QY 8 LLAAMVA-----GATATVPVPMVHPQPCACQIRWYPRSSYRATTVDCN 56
  |||
  |||
  |||
Db 13 LIAFWALGSGCHIOGDDPGASADAE---GPPCVACTS-----HDDYDDELVSFCS 61
  |||
  |||
  |||
QY 57 DLEFLTAVPALPAGTQTLTLLQNSIVRV-----DOSELGY 91
  |||
  |||
  |||
Db 62 SKNLTFLPDPIPVSTRALMDGNNLSIPSAFQNLSDIFLNLQSWLRLEPQALLG- 120
  |||
  |||
  |||
QY 92 LANLTLEDLSQNSFSQARDCDFALPQLLSLHEENQTRLE----- 133
  |||
  |||
  |||
Db 121 LQNLVYLHERNLRLNLAAGLFTHPSLASLSSNMLGRLEGLFQGLSHLMDLNLGNW 180
  |||
  |||
  |||
QY 134 -----DHSFAGLASIOELYLNHNOYRLAPRAFSGLSNLRHLNSNLRRAIDSWFEM 187
  |||
  |||
  |||
Db 181 SLVVLDPDYVQGLNHLHELVLNAGNKLTLYLPALFCGLRELDLSRNALRSVKANFVH 240
  |||
  |||
  |||
QY 188 LPNLLELITGKNKVDAL-----LDMN----- 208
  |||
  |||
  |||
Db 241 LPRLQKLYLDRNLITAVPAGAFGLGMKALRMLDLSHNRVAGLMEDFPGLGLHYRLAHN 300
  |||
  |||
  |||
QY 209 -----FRPLANRSVLVLAGMNLREISDYALSGLSLESISFYDNLAV----- 252
  |||
  |||
  |||
Db 301 AVALSRPRTFKDLHLEELQGLHNRLROGERTFEGIQLEVLTLINDQITEVRVGAFGSG 360
  |||
  |||
  |||
QY 253 -----PRRALEOVPEGLKFLDKNKNPLORVGPSPGANLHLKELGINNM 295
  |||
  |||
  |||
Db 361 LFNVAVMNLSGNCRLSLPRERVQGLDKLSLHLSCHLGHVLYLHTFPAGLSGLRRLFLRD- 419
  |||
  |||
  |||
QY 296 EEELVSDKALVNLPELTLDITNNPRLSFTIHPRAFNHLPOKETTMLNNALSAV----- 350
  |||
  |||
  |||
Db 420 NSISSIEEQLAGLSLELDELDTN-RLTHLPRQLFQGLGHEVLYLLSYNQLTTLSAEVL 478
  |||
  |||
  |||
QY 351 -----HQ-----QVESLPLNIQEVGLHGNPIR 372
  |||
  |||
  |||
Db 479 GELQRAFWDISHNHLLETLAEGLFSSLGVRVYLSLRNNSLOFSPQGLERLMDLANPMD 538
  |||
  |||
  |||
QY 373 CQCVR-----WAMATGTRVRIEP-----QSTLCAPPDLQRLPVEVPER 414
  |||
  |||
  |||
Db 539 CSCPLKALBDALQNPNGVVPFRVQTVCGSDQCPVYITNNITGACAPAVNSGLLDVSET 598
  |||
  |||
  |||
QY 415 EMTDHC 420
  |||
  |||
  |||
Db 599 HFV-HC 603

```

```

RESULT 5
ID AL5_MOUSE STANDARD; PRT; 603 AA.
AC P70389;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Insulin-like growth factor binding protein complex acid labile chain
  precursor (ALS).
GN IGFBP3 OR ALS OR ALBS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=96413591; PubMed=8816745;
RA Boisclair Y.R., Seto D., Hsieh S., Hurst K.R., Ooi G.T.;
RT "Organization and chromosomal localization of the gene encoding the
  mouse acid labile subunit of the insulin-like growth factor binding
  protein complex."
RL Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
CC -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
  CIRCULATING IGFs TO THE TISSUES.
CC -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
  IGF-1 OR IGF-TI AND IGFBP-3 (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U66900; AB017270.1; -.
DR MGD; MGI:107973; Igfbp3.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_cyp.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF01463; LRRCT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 4.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 9.
DR GlycoProtein; 1.
KW SIGNAL
FT 1 23
FT CHAIN
FT 24 603
FT REPEAT 52 73 LRR 1.
FT REPEAT 74 96 LRR 2.
FT REPEAT 98 120 LRR 3.
FT REPEAT 121 144 LRR 4.
FT REPEAT 146 168 LRR 5.
FT REPEAT 169 192 LRR 6.
FT REPEAT 194 216 LRR 7.
FT REPEAT 217 240 LRR 8.
FT REPEAT 242 264 LRR 9.
FT REPEAT 266 288 LRR 10.
FT REPEAT 289 312 LRR 11.
FT REPEAT 313 336 LRR 12.
FT REPEAT 338 360 LRR 13.
FT REPEAT 361 384 LRR 14.
FT REPEAT 386 408 LRR 15.

```


FT	REPEAT	765	788	LRR 21.
FT	REPEAT	790	812	LRR 22.
FT	REPEAT	813	836	LRR 23.
FT	REPEAT	838	861	LRR 24.
FT	DOMAIN	907	944	EGF-LIKE 1.
FT	DOMAIN	946	983	EGF-LIKE 2.
FT	DOMAIN	985	1022	EGF-LIKE 3.
FT	DOMAIN	1024	1062	EGF-LIKE 4.
FT	DOMAIN	1064	1100	EGF-LIKE 5.
FT	DOMAIN	1111	1149	EGF-LIKE 6.
FT	DOMAIN	1152	1325	LAMININ G-LIKE.
FT	DOMAIN	1353	1392	EGF-LIKE 7.
FT	DOMAIN	1409	1480	CTCK.
FT	CARBOHYD	111	111	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	207	207	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	357	357	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	435	435	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	783	783	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	788	788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	958	958	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	998	998	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1060	1060	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1159	1159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1243	1243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1292	1292	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	911	922	BY SIMILARITY.
FT	DISULFID	916	932	BY SIMILARITY.
FT	DISULFID	934	943	BY SIMILARITY.
FT	DISULFID	950	961	BY SIMILARITY.
FT	DISULFID	955	971	BY SIMILARITY.
FT	DISULFID	973	982	BY SIMILARITY.
FT	DISULFID	989	1001	BY SIMILARITY.
FT	DISULFID	995	1010	BY SIMILARITY.
FT	DISULFID	1012	1021	BY SIMILARITY.
FT	DISULFID	1028	1041	BY SIMILARITY.
FT	DISULFID	1035	1050	BY SIMILARITY.
FT	DISULFID	1052	1061	BY SIMILARITY.
FT	DISULFID	1068	1079	BY SIMILARITY.
FT	DISULFID	1073	1088	BY SIMILARITY.
FT	DISULFID	1090	1099	BY SIMILARITY.
FT	DISULFID	1115	1125	BY SIMILARITY.
FT	DISULFID	1120	1137	BY SIMILARITY.
FT	DISULFID	1139	1148	BY SIMILARITY.
FT	DISULFID	1357	1368	BY SIMILARITY.
FT	DISULFID	1362	1380	BY SIMILARITY.
FT	DISULFID	1382	1391	BY SIMILARITY.
FT	DISULFID	1409	1443	BY SIMILARITY.
FT	DISULFID	1423	1457	BY SIMILARITY.
FT	DISULFID	1434	1473	BY SIMILARITY.
FT	DISULFID	1438	1475	BY SIMILARITY.
FT	DISULFID	1442	1479	BY SIMILARITY.
FT	DISULFID	1394	1404	MISSING (IN SHORT ISOFORM).
SO	SEQUENCE	1480 AA;	165752 MW;	P9D5925FC170BC3 CRC64;

```

Query Match      8.7%  Score 324;  DB 1;  Length 1480;
Best Local Similarity 24.7%  Pred. No. 2,6e-15;
Matches 103;  Conservative 67;  Mismatches 169;  Indels 78;  Gaps 11;

QY  29  CPDCCACGLRPMWYTPRSSRYEATVYDNCNLFLLAAPPALPACTQTLILLOSNIIVRVDSE  88
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  519  CPAMCHC-----EGTYVDCITGRRLKEIPRIPIPLHTTELLLDNDELGRSSG  565

QY  89  LGLANLTELIDSONSFSDARDCEFFALPOLLSLHLEENQTLRLDEHSFAGIASLOEYL  148
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  566  L-----FCRLPLHYKLELTKRQNGIGEPNAEFGASHIQELQ  602

QY  149  NNNQLYRIAPRAFSGLSNLRLHLNSNLLRALDSKRFEMLPMEITMIGNKVYDALDMN  208
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  603  GEKKIKIEKNKMFGLJHOKLTNLIDNQISCPWGSFEHLNLSITSLIASNPFNC--NCH  660

QY  209  FRIANLRSLVLAGMNLR-----EISDVALGSLGLE-----  240

```

Db	661	LAMFAECYRKRSKSLINGGAARCAPSPKDYVQKDLPHSEFKCSSENSESGCLDGCYPCPST	720
Qy	241	-----SLSTFYDQGLARVPRALEQYPG-LKFLDILKNPDLQRYGPDGFANMLHKELGI	295
Db	721	CTGVVACSRNQDLKEIPR---GIPATSELTSELTSENEIQIHERIRHLSLSTRDLISN-	775
Qy	296	ELVSLIDKFAVLNYPETLKLDITNNPRLSFHPRAFHILPDMETLMIINNLSALHQOYV	355
Db	776	NQITLISNTYNTANLTKLSTLLISYN-KIQGLQRHALLSGLNLLRVSLHGRNISLPGSGF	834
Qy	356	ESLPNLQEVGHLGNEFRCDVYIRMANATGTRVRFIEPOSTICAEPPDLQRLPVREVP	412
Db	835	EDLKSLTFHIALGNSNPLYCDGLKWF-SPWIKLDVYEPGIARCAPEEOMKDKLLISTP	890
RESULT	8		
REV_HUMAN			
ID	GPV_HUMAN	STANDARD:	PRT: 560 AA.
AC	P40197;		
Dt	01-FEB-1995 (Rel. 31, Created)		
Dt	01-FEB-1995 (Rel. 31, Last sequence update)		
Dt	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Platelet glycoprotein V precursor (BPV) (CD42d).		
GN	GP5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KC	TISSUE=Lung;		
RX	MEDLINE=93391348; PubMed=7690959;		
RA	Hickey M.J., Hagen F.S., Yagil M., Roth G.J.;		
RT	"Human platelet glycoprotein V: characterization of the polypeptide		
RT	and the related Ib-V-IX receptor system of adhesive, leucine-rich		
RT	glycoproteins.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:8327-8331(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
KC	TISSUE=Platelet;		
RX	MEDLINE=94012616; PubMed=8407908;		
RA	Lanza F., Morales M., de la Salle C., Cazenave J.-P., Clemetson K.J.		
RA	Shimomura T., Phillips D.R.;		
RT	"Cloning and characterization of the gene encoding the human platelet		
RT	glycoprotein V. A member of the leucine-rich glycoprotein family		
RT	Cleaved during thrombin-induced platelet activation.";		
RL	J. Biol. Chem. 268:20801-20807(1993).		
RN	[3]		
RP	PARTIAL SEQUENCE.		
KC	TISSUE=Platelet;		
RX	MEDLINE=90775263; PubMed=2350580;		
RA	Shimomura T., Fujimura K., Maehama S., Takemoto M., Oda K.,		
RA	Fujimoto T., Oyama R., Suzuki M., Ichihara-Tanaka K., Titani K.,		
RA	Kuramoto A.;		
RT	"Rapid purification and characterization of human platelet		
RT	glycoprotein V: the amino acid sequence contains leucine-rich		
RT	repetitive modules as in glycoprotein Ib.";		
RL	Blood 75:2349-2356(1990).		
RN	[4]		
RP	PARTIAL SEQUENCE.		
KC	TISSUE=Platelet;		
RX	MEDLINE=90331220; PubMed=2327284;		
RA	Roth G.J., Church T.A., McMillen B.A., Williams S.A.;		
RT	"Human platelet glycoprotein V: a surface leucine-rich glycoprotein		
RT	related to adhesion.";		
RL	Biochem. Biophys. Res. Commun. 170:153-161(1990).		
CC	-I- FUNCTION: THE GPB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND		
CC	FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT		
CC	PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO		
CC	INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A		
CC	CRITICAL INITIATING EVENT IN HEMOSTASIS.		
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.		

[illegible]

INURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
 CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
 - SUBCELLULAR LOCATION: Type I membrane protein.
 - SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: Z69594; CAA93440.1; -
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_tyr.
 DR Pfam: PF00560; LRR; 14.
 DR Pfam: PF01463; LRRCT; 1.
 DR PRINTS: PR00019; LEUCRCHPT.
 DR SMART: SM00370; LRR; 2.
 DR SMART: SM00082; LRRCT; 1.
 DR SMART: SM00013; LRRNT; 1.
 DR SMART: SM00369; LRR_Tyr; 10.
 KW Platelet; Transmembrane; Glycoprotein; Blood coagulation;
 KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 567 PLATELET GLYCOPROTEIN V.
 FT DOMAIN 17 522 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 523 543 POTENTIAL.
 FT DOMAIN 544 567 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 73 96 LRR 1.
 FT REPEAT 97 120 LRR 2.
 FT REPEAT 122 144 LRR 3.
 FT REPEAT 145 168 LRR 4.
 FT REPEAT 169 192 LRR 5.
 FT REPEAT 194 216 LRR 6.
 FT REPEAT 217 240 LRR 7.
 FT REPEAT 241 264 LRR 8.
 FT REPEAT 266 288 LRR 9.
 FT REPEAT 289 312 LRR 10.
 FT REPEAT 314 337 LRR 11.
 FT REPEAT 338 361 LRR 12.
 FT REPEAT 362 385 LRR 13.
 FT REPEAT 387 409 LRR 14.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 567 AA; 63344 MW; CA10708ED03707F CRC64;

 Query Match 8.4%; Score 313.5; DB 1; Length 567;
 Best Local Similarity 24.0%; Pred. No. 3.8e-15;
 Matches 133; Conservative 62; Mismatches 226; Indels 133; Gaps 13;

 QY 4 LVAPLLAMVAGATATVYVPMHVPCCPGACQIRPWTPSSREATVVCNLEFLAV 63
 DB 1 MURSVLSAVLSLVGAQP-----PPCPRTCKGVK-----DAVCGSGSVNHI 43
 QY 64 PP-ALPAG-----TQTLLOSNISIVVDQSELGYANTLTEL 98
 DB 44 AELGLPTMLTILLFRMDRGVLOSHSFGMTVLOKIMLSDSHISAIDPGFTNDLVKIKTL 103
 QY 99 DLQSQSFSDARDQDFHALPOLSLHLEENQITRLHDSHFGALSLQELVYNHNOLYRIAP 158
 DB 104 RLTRNKISHLPRALDKWVLLLEQLFDHNAFLDLDDNLFKRLNLRLCLINONQLSFLPA 163

QY 159 RAESGLSNLRHLNLSNLRALIDSRRWEMLPNELIMIGKKVADILDMNPLANTSL 218
 DB 154 NLFSSGLKAKVDLSRRNLTHLPQGLGAQIKTKLLLYSNR---LMSLDSGLLANGAL 220
 QY 219 V---LAGNLRREISDYALEGLQSLSESYFYDNOQARVPRALEQVPKLFEDLNKNPLOR 275
 DB 221 TELRLERHNRSLAPGAVDSLGNLSTLISGNLLESIPRALFLVMSLTUTLTFENLEE 280
 QY 276 VGPQDFAMMLHLKELGNNMEELVSDKFAIVNIPELTKDITNNPRLSTHPRAFHNP 335
 DB 281 LPEVLFGEMAGLRELLWNG--THLRTLPAAAFRRNISGQITGLTRNPPLSLPQGMFGLT 339
 QY 336 QMETLMLNNALSLHQQTVESLPLQEVGLHGNPR----- 372
 DB 340 ELRLVAVHTNLEELPEALNGRLQVSLRNHRLALPRTLFRNLSSLVTVQLEHNL 399
 QY 373 -----CDC---VIRMANATGTRVRYRTEPOSTLCAEPDDL 403
 DB 400 KTLPGDVFAALPOLTRVLLGHNPWLCDGLMPFLQWLRHHELLGRDEPQ--CNGPESR 457
 QY 404 QRLPV-----REVPFRMTDHLPLISPPSPPSIQ-----VASGESMWL 443
 DB 458 ASLTFWELLQGDQWCPSSRGILPDPPTENALKAPDPQORNSSQSMAVOLVARGES--- 514
 QY 444 HCRALAEPEPEIYW 457
 DB 515 -----PDNRFYW 521

 RESULT 11
 ID CHAD_BOVIN STANDARD; PRT; 361 AA.
 AC Q27972;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Chondroaderlin precursor (Cartilage leucine-rich protein) (38 kDa bone
 DE protein).
 GN CHAD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Cartilage;
 RX MEDLINE=94342341; PubMed=8063792;
 RA Name P.J., Sommarin Y., Boynton R.E., Helnegaard D.;
 RT "The structure of a 38-kDa leucine-rich protein (chondroaderlin)
 RT isolated from bovine cartilage.";
 RL J. Biol. Chem. 269:21547-21554(1994).
 RN [2]
 RP SEQUENCE OF 25-55 AND 77-97.
 RC TISSUE=bone;
 RX MEDLINE=95113864; PubMed=7814406;
 RA Hu B., Coulson L., Moyer B., Price P.A.;
 RT "Isolation and molecular cloning of a novel bone phosphoprotein
 RT related in sequence to the cysstatin family of thiol protease
 RT inhibitors.";
 RL J. Biol. Chem. 270:431-436(1995).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: U08018; AAA21330.1; -
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.

```

DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR; 10.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 5.
KW Repeat; Signal.
FT SIGNAL 1 24 OR 23 (IN SOME ISOFORM(S)).
FT CHAIN 25 361 CHONDRODHERIN.
FT CHAIN 25 352 CHONDRODHERIN, MINOR FORM.
FT DOMAIN 79 102 10 x 24 AA LEUCINE-RICH TANDEM REPEATS.
FT REPEAT 103 102 1.
FT REPEAT 127 126 2.
FT REPEAT 151 174 3.
FT REPEAT 151 174 4.
FT REPEAT 175 198 5.
FT REPEAT 199 222 6.
FT REPEAT 223 246 7.
FT REPEAT 248 271 8.
FT REPEAT 272 293 9.
FT REPEAT 294 317 10.
FT DISULFID 306 348
FT DISULFID 308 328
FT CONFLICT 25 25 C -> Y (IN REF. 2).
FT CONFLICT 29 29 C -> W (IN REF. 2).
FT CONFLICT 31 31 C -> H (IN REF. 2).
FT CONFLICT 40 40 C -> L (IN REF. 2).
FT CONFLICT 52 52 S -> R (IN REF. 2).
SQ SEQUENCE 361 AA; 40884 MW; DA79DC98AD3DD1F8 CRC64;

```

```

Query Match 8.0%; Score 298.5; DB 1; Length 361;
Best Local Similarity 25.4%; Pred. No. 2,4e-14;
Matches 106; Conservative 64; Mismatches 152; Indels 95; Gaps 13;

```

```

QY 8 LILAWAGATATPVVWVWVPCPOACQIRPYTPRSSYREATVWDCNDFLTAVPRAL 67
DB 6 LILXSLGLIAS--LILPALACPCQNC-----HSDIQVTCXKVGLOKIPKY- 51
QY 68 PACTQTLLLOSNISIVWDSGLYLANLTELDSQNSFSDARDCDFHALPQLLSHLEEN 127
DB 52 -----SEKTKL-----LNLQRNNPVLTATNSFRAPNLYSLHLQHC 87
QY 128 QLTRLDEHSFAGIASIQELYNHNOYLRIAPAFSGISNLRLHLNSNLRAIDSWFEM 187
DB 88 QIREVAGAFRGKQILYILYLSHNDIRVLRAGAFDDITELTYLYLDHNRVTELPRGLSP 147
QY 188 LPLLEITMIGKNVDVAILDMNFRPLANLSVLAGNLRISYALEGLQSLLESISFYDN 247
DB 148 LVNLFILQLNNKIRLRSGAFGADLRILYSENSLSLOGALDDVENLAKFTLDN 207
QY 248 QLARVPRALEOVYGLKFLDLNKNPLQRYGPGDFANM-LHLKELGINNEELVSDK-- 304
DB 208 QLSYSPALSKLRVVEELKLSHPPLKSPDPAFQSGFRGLTETLMDN-----TNLEKFS 263
QY 305 -ALVNPBELTKDLITNNPRLSFTHPRAFHHL-----QMETLMMNNLSALHQQTVS 357
DB 264 GARLGVTTLKHVLENN-----RLHQLPSNFPDSELTETLTN----- 301
QY 358 LPLNQLVEGLGNPIKDCVI-----RWANATGTPRVREIEPOSTLCAEPDQLRPLVRE 410
DB 302 -----NPKKCTCOQLRGLRWLEAKTSR-----PDAT-CASPAKFRQGHIND 341

```

```

RESULT 12
CBP8 HUMAN STANDARD; PRT; 536 AA.
ID CBP8_HUMAN
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Carboxypeptidase N 83 kDa chain (Carboxypeptidase N regulatory
DE subunit) (Fragment).
GN CPN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90094386; PubMed=2378615;
RA Tan F., Weerasinghe D.K., Skidgel R.A., Tamei H., Kaul R.K.,
RA Robinson I.B., Schilling J.W., Erdos E.G.;
RT "The deduced protein sequence of the human carboxypeptidase N high
RT molecular weight subunit reveals the presence of leucine-rich tandem
RT repeats."
RL J. Biol. Chem. 265:13-19(1990).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=88309120; PubMed=3408501;
RA Skidgel R.A., Bennett C.D., Schilling J.W., Tan F., Weerasinghe D.K.,
RA Erdos E.G.;
RT "Amino acid sequence of the N-terminus and selected tryptic peptides
RT of the active subunit of human plasma carboxypeptidase N: comparison
RT with other carboxypeptidases."
RL Biochem. Biophys. Res. Commun. 154:1323-1329(1988).
CC -1- FUNCTION: THE 83 kDa SUBUNIT BINDS AND STABILIZES THE CATALYTIC
CC SUBUNIT AT 37 DEGREES CELSIUS AND KEEPS IT IN CIRCULATION. UNDER
CC SOME CIRCUMSTANCES IT MAY BE AN ALLOSTERIC MODIFIER OF THE
CC CATALYTIC SUBUNIT.
CC -1- SUBUNIT: TETRAMER OF TWO CATALYTIC CHAINS AND TWO GLYCOSYLATED
CC INACTIVE CHAINS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGION (POTENTIAL).
CC -1- PTM: WHETHER OR NOT ANY CYS RESIDUES PARTICIPATE IN INTRACHAIN
CC BONDS IS UNKNOWN, BUT THEY DO NOT FORM INTERCHAIN DISULFIDE BONDS
CC WITH THE 50 kDa CATALYTIC SUBUNIT.
CC -1- DISEASE: A COMPLETE ABSENCE OF THE ENZYME IS NOT CONSIDERED TO BE
CC COMPATIBLE WITH LIFE.
CC -1- SIMILARITY: CONTAINS 13 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J05158; AAA51921.1; -.
DR PIR; A34901; A34901.
DR MIM; 603104; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00560; LRR; 11.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 10.
KW Repeat; Leucine-rich repeat; Glycoprotein.
FT NON_TER 1 1
FT REPEAT 51 74 LRR 1.
FT REPEAT 75 98 LRR 2.
FT REPEAT 99 122 LRR 3.
FT REPEAT 123 146 LRR 4.
FT REPEAT 148 170 LRR 5.
FT REPEAT 171 194 LRR 6.
FT REPEAT 196 218 LRR 7.
FT REPEAT 220 242 LRR 8.
FT REPEAT 243 266 LRR 9.
FT REPEAT 268 290 LRR 10.
FT REPEAT 291 314 LRR 11.
FT REPEAT 316 338 LRR 12.
FT REPEAT 339 362 LRR 13.

```

CC		pancreas. Also detected in CD11c+ immature dendritic cells.
CC		Only expressed in dendritic cells and not in other leukocytes,
CC		including monocyte precursors.
CC	-1-	SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.
CC	-1-	SIMILARITY: CONTAINS 1 TIR DOMAIN.
CC	-1-	SIMILARITY: CONTAINS 22 LEUCINE-RICH REPEATS (LRR).
CC		
CC		This Swiss-Prot entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isdb.ch/announce/
CC		or send an email to license@isdb-sib.ch).
CC		-----
DR	EMBL:	U88879; AAC34134.1; .
DR	MIM:	603029; .
DR	InterPro:	IPR001611; LRR.
DR	InterPro:	IPR000483; LRR_Cterm.
DR	InterPro:	IPR000372; LRR_Nterm.
DR	InterPro:	IPR003592; LRR_out.
DR	InterPro:	IPR003591; LRR_Typ.
DR	InterPro:	IPR000157; TIR.
DR	Pfam:	PF00560; LRR_18.
DR	Pfam:	PF01463; LRRCt; 1.
DR	Pfam:	PF01582; TIR; 1.
DR	PRINTS:	PR00019; LEURICHRPT.
DR	SMART:	SM00370; LRR_5.
DR	SMART:	SM00082; LRRCt; 1.
DR	SMART:	SM00013; LRNT; 1.
DR	SMART:	SM00369; LRR_Typ; 2.
DR	SMART:	SM00255; TIR; 1.
DR	PROSITE:	PS50104; TIR; 1.
KW		Receptor; Immune response; Inflammatory response; Signal;
KW		Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein.
FT	SIGNAL	1 21 POTENTIAL.
FT	CHAIN	22 904 TOLL-LIKE RECEPTOR 3.
FT	DOMAIN	22 704 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	705 725 POTENTIAL.
FT	DOMAIN	726 904 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50 73 LRR 1.
FT	REPEAT	75 97 LRR 2.
FT	REPEAT	98 121 LRR 3.
FT	REPEAT	123 145 LRR 4.
FT	REPEAT	147 169 LRR 5.
FT	REPEAT	170 193 LRR 6.
FT	REPEAT	196 219 LRR 7.
FT	REPEAT	247 270 LRR 8.
FT	REPEAT	273 296 LRR 9.
FT	REPEAT	298 320 LRR 10.
FT	REPEAT	329 377 LRR 11.
FT	REPEAT	379 403 LRR 12.
FT	REPEAT	406 429 LRR 13.
FT	REPEAT	431 454 LRR 14.
FT	REPEAT	455 478 LRR 15.
FT	REPEAT	480 504 LRR 16.
FT	REPEAT	505 528 LRR 17.
FT	REPEAT	530 552 LRR 18.
FT	REPEAT	561 584 LRR 19.
FT	REPEAT	586 608 LRR 20.
FT	REPEAT	610 632 LRR 21.
FT	REPEAT	634 659 LRR 22.
FT	TIR.	
FT	CARBOHYD	52 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	57 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	70 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	124 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	196 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	247 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	252 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	255 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	262 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	275 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	291 N-LINKED (GLCNAC...) (POTENTIAL).

```

FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 662 662 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 904 AA; 103828 MW; 034E05ECA7A4D2P7 CRC64;

Query Match 7.7%; Score 287; DB 1; Length 904;
Best Local Similarity 26.1%; Pred. No. 6.1e-13;
Matches 95; Conservative 65; Mismatches 142; Indels 62; Gaps 12;

QY 74 LILQNSIYRVDSSELYGANTLDELDSQNSFSDARDCDFHALPOLLS-----LHLENOQ 128
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 360 LNMEDNDIGIKSNMGTGINIKYLSL--NSTSLRTLTNEFFVSLAIPHILMLTKNK 418
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 129 LRLLEDHSPAGLASLOELYLNHOQ--YRIAPRASGLSNLRLHLNSILRAIDSRWEP 187
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 419 ISKIESDAFSGHLEVLIDGLINEIGELTGOEWKLENIPEIYLSYNKYQLQTRNSPAL 478
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 188 LPLLETLI---IGKNKYVDALIDMNPRLANLSIVLAGNLEISDYALEGQSLESSEF 244
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 479 VPSLQRLMLRRVALKNVDS--SPSPFQPLRNLTLLDSNNNNININDMLEGLEKLEILD 537
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 245 YDQQLARVPRA-----LEQVPLKFLDLINKNPLQVCPGDFANMLHLK--ELGLNN 294
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 538 QHNHNLARKKHNKPGPIYFIKGLSHLHLNLESNGFDIPEVFEVDLELTIIDGLNN 597
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 295 MEELVSIKFAVNLPELTKLDTNNPRLSFIHRAFHLLPOMETLMNNALSALHOOT 354
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 598 LNTLPA-----SVENNQVSLKSLNLOKNLITSEKRV 629
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 355 V-ESLNLQVGLHGNPIRQDC-----VIRMANATGRVRETEPOST---LCAEPDQLR 405
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 630 ECGAFRNLTLELDRFNPDCSTCESTAFVAINETHINI-----PELSHYLCNTPPHYHG 685
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 406 LPVR 409
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 686 FPVR 689

RESULT 14
GAP_HUMAN STANDARD; PRT; 662 AA.
ID GAP_HUMAN 014392;
AC 014392;
DT 01-NOV-1997 (Rel. 35, Last sequence created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE GAP protein precursor (garpin) (glycoprotein A repetitions
DE predominant).
DE
GN GAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94235567; PubMed=8180135;
RA Ollendorff V., Noguchi T., Delapeyriere O., Birnbaum D.;
RT "The GAP gene encodes a new member of the family of leucine-rich
RT repeat-containing proteins."
RL Cell Growth Differ. 5:213-219(1994).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 22 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; 224680; CAA80847.1; -.
DR MIM; 137207; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_Cyp.
DR Pfam; PF01462; LRRNT; 1.
DR Pfam; PF00560; LRR; 17.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_Typ; 2.
DR SMART; SM00369; LRR_Typ; 2.
KW Glycoprotein; Leucine-rich repeat; Repeat; Transmembrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 662
FT DOMAIN 20 627
FT TRANSMEM 628 648
FT DOMAIN 649 662
FT DOMAIN 8 31
FT REPEAT 48 71
FT REPEAT 72 95
FT REPEAT 97 122
FT REPEAT 123 147
FT REPEAT 148 171
FT REPEAT 172 195
FT REPEAT 197 217
FT REPEAT 218 240
FT REPEAT 242 264
FT REPEAT 265 286
FT REPEAT 314 337
FT REPEAT 339 361
FT REPEAT 362 385
FT REPEAT 386 408
FT REPEAT 410 432
FT REPEAT 442 465
FT REPEAT 467 488
FT REPEAT 490 513
FT REPEAT 514 538
FT REPEAT 540 558
FT REPEAT 560 583
FT REPEAT 603 623
FT CARBOHYD 271 271
FT CARBOHYD 308 308
FT CARBOHYD 345 345
FT CARBOHYD 545 545
SQ SEQUENCE 662 AA; 71978 MW; 700B93ACDCE25960 CRC64;

Query Match 7.5%; Score 280.5; DB 1; Length 662;
Best Local Similarity 22.6%; Pred. No. 1.1e-12;
Matches 130; Conservative 50; Mismatches 159; Indels 235; Gaps 14;

QY 53 VDCNDLFTAVPAPLACTQTLLQNSIVRDSSELYGANTLDELDSQNSFSDARDCD 112
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 33 VSCQVLGILQVPSVLPPTETLDSGNQLRSILASPLGFTYALRHLDLSTNEISFLPGA 92
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 113 FHAL-----LDM-----PQL 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 93 FQALTHLEHLSTAHNRLAMATASAGLGPLPRYTSIDLSGNSISGLLERLGAEVSLH 152
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 121 SLHLEENQTLREDSFAGLASLOELYLNHOQYRIAPRASGLSNLRLHLNSILRAI 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 153 TSLAENSTLRLTRTFTEPMDLPEOLDLHSVNLMDIEDGAGEGLPRLTHLMSRNSLTCT 212
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 181 DSRWFEMLPNLEILMIGNKYDAI-----LDM----- 207
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 213 SD---FSIQQLRVLDLSCNSTEAFQASQPAEQTLWLDLREKKLHFPLALPRLIY 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 208 -----NRPRLANRSIVLAGNIRE 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 270 LMLSNNLRLPTGPQDSKGIHAPSEGWSALPLASGNSAGSRPLSQLINDLSYNEIEL 329
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 228 ISDYALEGQSLESSEFYDNLQARVPRALDEVPOLKFLDLINKNPLQK----- 275
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

```

Db 330 IPDSFELHLSLCLNLSRRCLETRFEARRLGSLPCLMLDLSHNALETTLELGARALGSLR 389
QY 276 -----VGPDPANMLHLKEI-----GNNMEELY 299
Db 390 TLLQGNALNDLPYRTANLASLQRLNLQGNRVSPCGPDEPGSGCVAFSGITSLRSLX 449
QY 300 SLDK-----FALVNLPELTJKTLDITNNP-----321
Db 450 LVDMIEILLRAGAFLEHFP-LTEILDLSNPGLEVATGALGSLASLEVLALQNGMLVQY 508
QY 322 -----RLSFTHPAFHHP-----QMETLMLNNNLSALHQQVESL-PNIQEWG 365
Db 509 DLPCFLCKRLNLAEKR-LSHLPATWQAVSLLEVLDLRNNSFSLPGSAMGLETSIRRLX 567
QY 366 LHGNPDRCDVIRW--ANATGTRVRFLEPSTIC 397
Db 568 LQGNPLSC-CGNGLMAQLHQGVADYADODLIC 600

RESULT 15
CHAO_TRICA STANDARD; PRT; 782 AA.
AC P82963;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaoptin (Photoreceptor cell-specific membrane protein) (Fragment).
GN CHP.
OS Tribolium castaneum (Red flour beetle).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OX Cuculiforma; Tenebrionidae; Tribolium.
ON NCBI_TaxID=7070;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-GA-1;
RA Nie W., Stronach B., Panganiban G., Brown S., Denell R.;
RT "Molecular characterization of Tc1abial and the 3' end of the
RT Tribolium homeotic complex.";
RL Submitted (NOV-2000) to the EMBL/Genbank/DDBJ databases.
CC -1- FUNCTION: REQUIRED FOR PHOTORECEPTOR CELL MORPHOGENESIS. MEDIATES
CC HOMOPHILIC CELLULAR ADHESION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR SURFACE OF R-CELL PLASMA
CC MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAOPTIN FAMILY.
CC -1- SIMILARITY: CONTAINS AT LEAST 21 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF322227; AK01654.1; --
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR Pfam: PF00560; LRR; 19.
DR PRINTS: PR000019; LEURICHRPT.
DR SMART: SM00370; LRR; 6.
DR SMART: SM00369; LRR_typ; 14.
KW Glycoprotein; Membrane; Repeat; Leucine-rich repeat; Vision.
FT NON_TER 1
FT REPEAT 43 66 LRR 1.
FT REPEAT 67 90 LRR 2.
FT REPEAT 93 116 LRR 3.
FT REPEAT 117 140 LRR 4.
FT REPEAT 141 164 LRR 5.
FT REPEAT 165 188 LRR 6.
FT REPEAT 224 247 LRR 7.

```

```

FT REPEAT 249 272 LRR 8.
FT REPEAT 273 296 LRR 9.
FT REPEAT 297 320 LRR 10.
FT REPEAT 321 343 LRR 11.
FT REPEAT 344 367 LRR 12.
FT REPEAT 370 391 LRR 13.
FT REPEAT 395 418 LRR 14.
FT REPEAT 419 444 LRR 15.
FT REPEAT 446 469 LRR 16.
FT REPEAT 491 513 LRR 17.
FT REPEAT 514 537 LRR 18.
FT REPEAT 539 562 LRR 19.
FT REPEAT 615 638 LRR 20.
FT REPEAT 640 663 LRR 21.
FT CARBOHYD 196 196 LRR 21.
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 782 AA; 88561 MW; 311460B2D4527917 CRC64;

```

Query Match 7.5%; Score 279; DB 1; Length 782;
 Best Local Similarity 24.8%; Pred. No. 1,9e-12;
 Matches 145; Conservative 94; Mismatches 233; Indels 112; Gaps 23;

```

QY 47 YREATVVCNDLEFLT-AVPPALPAGTQTLILQNSIVVDGSEGLYANLELDLSSNSF 105
Db 5 HRKQFVYIGSGSLTKLTKLTHL-----SSVQNFPSDAKILNREELDLSSNRL 54
QY 106 SDARDCDFHALPOLSLHLEENQTLREDSFAG--LASLQELYNHNOLYRIAPRAFSG 163
Db 55 RNVDPNSFHFRLSKLVHLQNTIEMIHRGTFQGDHHRDLTEVEFSNVSANVQGHFTF 114
QY 164 LSNLRRLNLSNLRAIDSRFEMLPNLEITMIGKNVDALDMNFRPLANBSLVLAGM 223
Db 115 LIQLEQIHLDNRJESLERRAFMNILSKLRNLNGNATATAYTFQNPDELEDLDAYN 174
QY 224 NLREISDYALBGLQSLLESIFYDN-----QLARVPRRALEQVPGI-----KFLDLNK 270
Db 175 SSSLDENFIQVGSIG--MFHYMNSHKILNLVVAASVPEQDTGLGGLQNIKVDLSF 232
QY 271 NPLQRVGQGDPAAM-LHLKEGLNNMEELYSIDKFAVLNPELTJKTLDITNNP--RLSFIH 327
Db 233 NNITSVAKQFFRPVELSLMOLYIGH-NKLINATKDLFGNMPHLOVLDLSHNSLYEIDF-- 289
QY 328 PRAFHNLPEOMETLMNNNALSLALHQQVESLPNIQEWGLHGNPTRC--DCVIRMANATGT 385
Db 290 -DTRNRKIKLQWLDTSHNRKISEINDLFRFLGNLRKYDFSHNRKRSKLPDLFR---ETG- 344
QY 386 RVRFLEPOSTLCABPDLQPLVREVPFRMTDCLPLISPRSFPSLQVAGSGESWYLHC 445
Db 345 -----LERLDVSHNLGK-----LPL-----TSLSLASQQLT----- 371
QY 446 RALAPPEPEIYVTPAGLRILTPAHAGR--RYRYVPESTLELRKYTAEBEAGLYCV---AQ 500
Db 372 -----SELDLSWNSISSL---SHGQLARFKCLSWLDSYNRKLGQIDAQTFGIDPRLAS 422
QY 501 NLVGADTKTVSVVYGRALQPGRDEGGLRFAQVETHPHYHLLSW---PPPTVSTNLT 557
Db 423 LNLGHNSQILFLEINGLSF-----QGLRYTLHLNLDVNSISQVPAISTPLSLSLDA 474
QY 558 WSG-----ASSLR--GGATAPLARLPGSTSYNITRLL 588
Db 475 FNSLPTVALEVAGNISSLRYLNLNDYNDLSAVPIVTHSLTFLRHL 518

```

Sat Aug 31 14:55:23 2002

us-09-905-088a-245.rsp

Page 16

Job time: 263 sec

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2002, 14:36:13 ; Search time 34.24 Seconds

(without alignments)
3602.377 Million cell updates/sec

Title: US-09-905-088a-245

Sequence: 1 MRLVAPFLIAWAGATATV.....RKLPSSSEGTLLPILSONS 713

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteint:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	1695	45.4	730	4 Q9P231	Q9P231 homo sapien
2	1687	45.2	716	11 Q61809	Q61809 mus musculu
3	1656.5	44.4	705	4 Q43377	Q43377 homo sapien
4	1656.5	44.4	708	4 Q9H3W5	Q9H3W5 homo sapien
5	1649.5	44.2	707	11 Q9ESY6	Q9ESY6 ratuS norv
6	1648.5	44.2	707	11 P97860	P97860 mus musculu
7	1625.5	43.6	718	13 Q73675	Q73675 xenopus lae
8	875.5	23.5	431	4 Q9H8V1	Q9H8V1 homo sapien
9	755.5	20.2	273	4 Q9HUY4	Q9HUY4 homo sapien
10	494.5	13.3	614	11 Q9D1T0	Q9D1T0 mus musculu
11	492.5	13.2	614	4 Q96F55	Q96F55 homo sapien
12	490	13.1	614	6 Q9N008	Q9N008 macaca fasc
13	468.5	12.6	606	4 Q9BZ20	Q9BZ20 homo sapien
14	442	11.8	1091	11 P70193	P70193 mus musculu
15	438.5	11.7	640	4 Q9HCJ2	Q9HCJ2 homo sapien
16	432.5	11.6	540	5 Q9VU53	Q9VU53 drosophila

17	431	11.5	1094	4 Q9BVB8	Q9BVB8 homo sapien
18	430	11.5	1093	4 Q9BVB1	Q9BVB1 homo sapien
19	427.5	11.5	532	5 Q96671	Q96671 drosophila
20	420.5	11.3	628	4 Q9BTN0	Q9BTN0 homo sapien
21	407	10.9	733	5 Q24250	Q24250 drosophila
22	407	10.9	737	5 Q9VU51	Q9VU51 drosophila
23	406.5	10.9	719	4 Q96NT6	Q96NT6 homo sapien
24	401.5	10.8	737	5 Q965M3	Q965M3 caenorhabdi
25	401.5	10.8	881	5 Q965M2	Q965M2 caenorhabdi
26	401.5	10.8	1447	5 Q16779	Q16779 caenorhabdi
27	399.5	10.7	811	4 Q75139	Q75139 homo sapien
28	399	10.7	78	11 Q61974	Q61974 mus musculu
29	398.5	10.7	789	6 Q9BE71	Q9BE71 macaca fasc
30	395.5	10.6	673	11 Q9C2T5	Q9C2T5 mus musculu
31	392.5	10.5	792	4 Q9UDT7	Q9UDT7 homo sapien
32	388.5	10.4	832	4 Q9ULH4	Q9ULH4 homo sapien
33	386	10.3	492	11 Q99KT6	Q99KT6 mus musculu
34	386	10.3	653	4 Q9HBL1	Q9HBL1 homo sapien
35	385	10.3	788	11 Q9CYK3	Q9CYK3 mus musculu
36	378	10.1	649	4 Q96A85	Q96A85 homo sapien
37	377	10.1	1065	4 Q94898	Q94898 homo sapien
38	374.5	10.0	700	4 Q9P244	Q9P244 homo sapien
39	368	9.9	907	11 Q921P4	Q921P4 mus musculu
40	361.5	9.7	542	5 Q9N4G6	Q9N4G6 caenorhabdi
41	361.5	9.7	809	11 Q9DBY4	Q9DBY4 mus musculu
42	361	9.7	1173	5 Q9V7J8	Q9V7J8 drosophila
43	360.5	9.7	1531	11 Q88279	Q88279 ratuS norv
44	360	9.6	473	11 Q99PI8	Q99PI8 mus musculu
45	360	9.6	907	4 Q75473	Q75473 homo sapien

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	730 AA.
Q9P231	Q9P231	Q9P231		
AC	Q9P231	Q9P231		
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	KIAA1497 PROTEIN (FRAGMENT).			
GN	KIAA1497.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20277482; PubMed=10819331;			
RA	Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;			
RT	Prediction of the coding sequences of unidentified human			
RT	genes. XVII. The complete sequences of 100 new cDNA clones from brain			
RT	and which code for large proteins in vitro.;			
RL	DNA Res. 7:143-150(2000).			
DR	EMBL: AB040930; BAA96021.1; -			
DR	InterPro: IPR003961; FN.III.			
DR	InterPro: IPR003598; Ig_C2.			
DR	InterPro: IPR003006; Ig_MHC.			
DR	InterPro: IPR001611; LRR.			
DR	InterPro: IPR000483; LRR_Cterm.			
DR	InterPro: IPR000372; LRR_Nterm.			
DR	InterPro: IPR003592; LRR_out.			
DR	InterPro: IPR003591; LRR_Typ.			
DR	Pfam: PF00047; Ig_1.			
DR	Pfam: PF00560; LRR_9.			
DR	Pfam: PF01463; LRRCT_1.			
DR	Pfam: PF01462; LRRNT_1.			
DR	SMART: SM00060; FN3_1.			
DR	SMART: SM00408; IGC2_1.			
DR	SMART: SM00370; LRR_4.			
DR	SMART: SM00082; LRRCT_1.			
DR	SMART: SM00013; LRRNT_1.			


```

DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PLACE1006239 PROTEIN (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Iisogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirooka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuho Y., Kanehori K.;
RT "NDO human cDNA sequencing project.";
RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001991; BAB92025.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR00372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00560; LRR; 6.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 2.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_typ; 1.
FT NON_TER 273
FT SEQUENCE 273 AA; 30832 MW; B59F070A6D7D11FF CRC64;

Query Match 20.2%; Score 755.5; DB 4; Length 273;
Best Local Similarity 58.8%; Pred. No. 1e-45;
Matches 151; Conservative 39; Mismatches 66; Indels 1; Gaps 1;

QY 17 TATVAVVWVHPVPCQACQIRPWTYPRSSYREATYVDCNDLEFLAVPALPAGTQTLT 76
DB 17 TTVQAVVKKVDCPRLCCEIRPWTYPRSSYREATYVDCNDLEFLAVPALPAGTQTLT 76
QY 77 QSNSTVAVVQSELGYLANITFLDLSQNSFSDARDCEHALPOLISLHLENOJLRLEDS 136
DB 77 QTNNTAKLEYS-TDEPVNLTGLDLSQNSFSDARDCEHALPOLISLHLENOJLRLEDS 135
QY 137 FAGLASLOELYNHOLYRIAPRAFSGSLRLHLNSNLRAIDSQWFMPLNLELMT 196
DB 136 LSELNSLOELYNHOLYRIAPRAFSGSLRLHLNSNLRAIDSQWFMPLNLELMT 195
QY 197 GKNVDAIDLNNFRPLANLRSVLAGMNLREISDYALEGLQSLSESYNOJLARVPRRA 256
DB 196 GNPPIIRIKDNKFRPLNLSVLAGINTLEIPDNALVGLLENESISFYDNRIKVPYHVA 255
QY 257 LEQVGLKFLDLNKNPL 273
DB 256 LQKVYNLKLFLDLNKNPI 272

RESULT 10
Q9D1T0 PRELIMINARY; PRT; 614 AA.
AC Q9D1T0:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHD LIBRARY,
DE CLOVE:4930471K13, FULL INSERT SEQUENCE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shnagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK027262; BAB32403.1; -.
DR HSSP; P23945; 1XUN.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_c2.
DR InterPro; IPR003606; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00560; LRR; 9.
DR Pfam; PF01462; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR SMART; SM00409; Ig; 1.
DR SMART; SM00408; IgC2; 1.
DR SMART; SM00410; Ig_Like; 1.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_typ; 9.
KW Immunoglobulin domain.
SQ SEQUENCE 614 AA; 69100 MW; 41CFE40C21335681 CRC64;

Query Match 13.3%; Score 494.5; DB 11; Length 614;
Best Local Similarity 25.6%; Pred. No. 2.9e-29;
Matches 173; Conservative 97; Mismatches 251; Indels 155; Gaps 18;

QY 1 MRLVAPILLAM-----VAGATVAVVWVHPVPCQACQIRPWTYPRSSYREA 50
DB 6 MRSMPSPILLACWOPILLVLGLSVLSGSAT-----GCPRECECS-----AAD 46
QY 51 TTVDCNDLEFLAVPALPAGTQTLTQNSIVRYDQSELGYLANITFLDLSQNSFSDARD 110
DB 47 RAVLCHRRKRYAVPDEGIPETRLIDLQKNKIKTLNODERFSFRLLELENEYAVAPER 106
QY 111 CDFH-----ALPQLSLHLENOJLRLEDSFAGIASLOEL 146
DB 107 GAFNLFMLRTILGRLSNRLKILPLGVFTGLSLNLFKLDISENKKIIVILIDYFQOLYMKSL 166
QY 147 YLNNHOLYRIAPRAFSGSLRLHLNSNLRAIDSQWFMPLNLELMTGCKVVAIID 206
DB 167 EVDGDDLYIISHRAFSGLNSLEQLTLKCNLTSTPEALSHLGLLYLRRLRNINVAID 226
QY 207 MNFRPLANLRSVLAGM-NLEISDYALEGLQSLSESYNOJLARVPRRALQVGLKF 265
DB 227 YSFKRLYLKVLKLEISHWYIDTTPNCLYGL-NLTSISITHCNLAVPYLAVYHLVYLF 285
QY 266 LDKNKPIQRYGPDFAFMMLHLKELGLNNEELVSDKFLVYMLPELTAKDITNNRSLF 325
DB 286 LNLSTNPIGTLE--GSMH-----ELRLQELQV-----GGQLAV 319

```

```

QY 326 IHPRAFHHLPEMETLMLNNALSALHQOTVESLPNLOEVGLHGNPIRDCCVIRMANATGT 385
D 320 VEPYAFRLNLYRLVNLVSGNQLTLEESVFSVGNLETLILDSNPLACDCRLMWFRRRW 379
QY 386 RVRFIEPOSTLCAEPDLOLRVREVPFREMTHCLPLISPRSF-----PPLSQ 434
D 380 RLNFNRQOPT-CATPEFVQGEKDFP-----DVLFPNFTCCRARIIDRKAQOVF 429
QY 435 VASGESMVLHCRALAEPEPEIYWTYPAGLRITPAHAGRRYRYPBGTLELRVTAEEAGL 494
D 430 VDEGHTVQFVCRADGPPAILMLSPR-KHLVSAKSNGLTYFPGDLEVRAYQVODNGT 488
QY 495 YTCVAQNLVGAADTKTVSVVVGRRALLQPGDEGQGLELRVQETHPHYHILLSWTPPNTYST 554
D 489 YLCIAANAGGNDM-----PAHLHVRSYSPDMPHOP 519
QY 555 NLTWSSASSLRGGATATALRPRGT--HSYNITRLQATE-----YMACLOVAEAD 603
D 520 NKTFAFISNQPGE--ANSTRATVPPFDIKTLIATMGFISFLGVVLCVLLF-- 574
QY 604 AHTOLACVWARTKEAT 619
D 575 -----LMSRGKNT 583

```

RESULT 11

```

Q96FE5 PRELIMINARY; PRT; 614 AA.
AC 096FE5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:17422).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE-BRAIN, AND ANAPLASTIC OLIGODENDROGLIOMA WITH 1P/19Q LOSS;
RA Strusberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011057; AAH1057.1; -.
SQ SEQUENCE 614 AA; 69145 MW; EFD967E3B716698D CRC64;

```

Query Match 13.2%; Score 492.5; DB 4; Length 614;
 Best Local Similarity 25.4%; Pred. No. 4,1e-29;
 Matches 172; Conservative 98; Mismatches 251; Indels 155; Gaps 18;

```

QY 1 MRLIVAPILLAW-----VAGATATVPVPMHVPCCPOCACQIRPWYPRSSYREA 50
D 6 VRMPSPILACWQPIILLVIGVLSGSAI-----GCPPECSC-----AGD 46
QY 51 TTVDCNDLFTAVPALPAGSTQTLQSNISYRVDSGLGYANTLTDLSONSFSQDAND 110
D 47 RAVLCGRKRVAVVEGIPETRLIDGKNRIKTLNODFASPFNLEINIVASVEP 106
QY 111 CDRH-----ALPOLSHLEENQTRFEDHSFAGLSLOEL 146
D 107 GARNNLFNLTGLRSNRKLPLGVFTGISNTFKLIDISENKTIVLLDYMFODLNLKSL 166
QY 147 YLHNHOLYRIAPAFSGLSNLRLHNSNLRAIDSRKWEMLPNTLEILMIGKKVDAIID 206
D 167 EVDGNDLVYISHAFSGLSNLEBUTLEKCLNITIPREALSHLGLVLLRHNINAIAD 226
QY 207 MNRPLANRSLVLAGM-NIREISDYALBGLQSLSESYDNOLAVRRALAEQVGLKF 265
D 227 YSRKRLVRLKVLLEISHMPYLDWTMPNCLVGL-NLTSLSITHCNLPAVYLAVHNLVLYRF 285
QY 266 LDKNRPLOVVGPGDFANMHLKELGLNNNEELVSDKFAVLVLPRLTKLIDITNNRSLSF 325

```

```

D 286 LNIStNPISITIE-----GSMH-----ELLRQLQELQV-----GGQLAV 319
QY 326 IHPRAFHHLPEMETLMLNNALSALHQOTVESLPNLOEVGLHGNPIRDCCVIRMANATGT 385
D 320 VEPYAFRLNLYRLVNLVSGNQLTLEESVFSVGNLETLILDSNPLACDCRLMWFRRRW 379
QY 386 RVRFIEPOSTLCAEPDLOLRVREVPFREMTHCLPLISPRSF-----PPLSQ 434
D 380 RLNFNRQOPT-CATPEFVQGEKDFP-----DVLFPNFTCCRARIIDRKAQOVF 429
QY 435 VASGESMVLHCRALAEPEPEIYWTYPAGLRITPAHAGRRYRYPBGTLELRVTAEEAGL 494
D 430 VDEGHTVQFVCRADGPPAILMLSPR-KHLVSAKSNGLTYFPGDLEVRAYQVODNGT 488
QY 495 YTCVAQNLVGAADTKTVSVVVGRRALLQPGDEGQGLELRVQETHPHYHILLSWTPPNTYST 554
D 489 YLCIAANAGGNDM-----PAHLHVRSYSPDMPHOP 519
QY 555 NLTWSSASSLRGGATATALRPRGT--HSYNITRLQATE-----YMACLOVAEAD 603
D 520 NKTFAFISNQPGE--ANSTRATVPPFDIKTLIATMGFISFLGVVLCVLLF-- 574
QY 604 AHTOLACVWARTKEAT 619
D 575 -----LMSRGKNT 583

```

RESULT 12

```

Q9N008 PRELIMINARY; PRT; 614 AA.
AC 09N008;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOHETICAL 69.2 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolpus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP TISSUE-CEREBELLUM CORTEX;
RA Oseada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046639; BAB03557.1; -.
DR HSSP; P23945; IXUN.
DR InterPro; IPR003598; Iq_C2.
DR InterPro; IPR003066; Iq_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_tyr.
DR Pfam; PF00047; Iq_1.
DR Pfam; PF00560; LRR_9.
DR Pfam; PF01463; LRRCT_1.
DR Pfam; PF01462; LRRNT_1.
DR SMART; SM00408; IGG2; 1.
DR SMART; SM00370; LRR_5.
DR SMART; SM00082; LRRCT_1.
DR SMART; SM00013; LRRNT_1.
DR SMART; SM00369; LRR_tyr_1.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 614 AA; 69187 MW; BA6CB8C7C933BE9A CRC64;

```

Query Match 13.1%; Score 490; DB 6; Length 614;
 Best Local Similarity 26.4%; Pred. No. 6,4e-29;
 Matches 160; Conservative 89; Mismatches 229; Indels 128; Gaps 16;

```

QY 1 MRLVAPLLAM-----VAGATATVPVWHPVPCPOACQIRPWTTPRSSYREA 50
Db 6 VASMSPLLACQPIILLVLSVLSGXT-----GCPRECS-----AOD 46
QY 51 TTVDCNDLFLAVPALPAGTQTLLOSNSIVRVDSGLYANLTDLSONSPSDARD 110
Db 47 RAVLCRRKRVAVPFGIPTETRLDGLGNRIKTLNODFASFPHELELMENIYSAVER 106
QY 111 CDFH-----ALQQLSLHEENQRLREHSHRAGLASLOEL 146
Db 107 GAFNNLFNLRTGLRSNRKLIPLGVFTGLSNLTKRLDISENKKIYLLDYMQDLYNLKSL 166
QY 147 YVNHQVRIAPRASGLSNLRHLNLSNRALDASRWEMLPNLEILMIGNKVDATLD 206
Db 167 EYGDNDLVYISHRAFSGLSLEQLTEKCNLTSTETALSHLGLIYLRHNLINATRD 226
QY 207 MNFRPLANLRSVLAM- NLREISDALEGLQSLSESYDNQALRVPRALQVPGIKF 265
Db 227 YSFKRLYRLKVLKISHWPYLDTPNCLYGL- NLTSITHCNLTAVPYLAVRHLVYLR 285
QY 266 DLNNKPLQVPGDPFANMLHKLGLNNMELVSDKFALVNPBELTKLDITNPRISF 325
Db 286 LNLSTNPISITIE-----GSLH-----ELRLQELQV-----GGQLAM 319
QY 326 IHPRAFHLPMQETLMLNNALSLALHOQVESLPLNLOEVLGNPDRCDVIRMANATGT 385
Db 320 VERPFAFRGLNLYRVLVNSQNLTLSESVFHSVGNLETLIDNSLADCDRLMWFRRRW 379
QY 386 RVRTEPOSTLCAEPDQRLPVRVPREMTDHLPLISPSRF-----PSLQ 434
Db 380 RLNFNQOPT-CAPEFVGKKEKDFP-----DYLLPWFYCRARIRIRKKAQYF 429
QY 435 VASGESMVLHCLALPEPEEYIWPVPAGLRLPAHAGRRYRYPGTELEIRVTEEGGL 494
Db 430 VEGHTVQFVCRADDPPEPAILMLSPR- KHLVSAKSNGLTYFPGTELEVARAOVDNGT 488
QY 495 YTCVAQNLVAGDT-----KTVSVVGRALLOPGDEGGLRLQVOE 535
Db 489 YTCIANMGNDMSRAHLHVRSYSPDWPQPKTKFAFIPN-----QPG-EGGANSTRATV 542
QY 536 THPYHT 541
Db 543 PEPFDI 548

```

RESULT 13

Q9B220 PRELIMINARY: PRT; 606 AA.

AC Q9B220;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, last annotation update)

DE BA48B23.1 (NEURONAL LEUCINE-RICH REPEAT PROTEIN) (CDNA FLJ1810 FIS, CLONE NT2R12009289, WEAKLY SIMILAR TO CARBOXYPEPTIDASE N 83 KDA DE CHAIN).

DE BA48B23.1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxId=9606;

OC [1]

RP SEQUENCE FROM N.A.

RA Babbage A.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y., Hotate T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., Sugiyama T., Irie R., Otsuki T., Sato H., Makamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

```

RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDD human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL353746; CAC22713.1; -.
DR EMBL: AK056372; BAB71167.1; -.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003591; LRR_typ.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00560; LRR_11.
DR Pfam: PF01463; LRRCT; 1.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00409; IG_1.
DR SMART: SM00408; IGc2; 1.
DR SMART: SM00410; IGc2; 1.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_typ; 10.
KW Immunoglobulin domain.
SQ
SEQUENCE 606 AA; 6806 MW; CB608E281B06B9D CRC64;

```

Query Match 12.6%; Score 468.5; DB 4; Length 606;

Best Local Similarity 27.0%; Pred. No. 2,7e-27;

Matches 140; Conservative 84; Mismatches 221; Indels 73; Gaps 9;

```

QY 27 VPCPOCACQIRPWTTPRSSYREATVDCNDLFLVAPPALPAGTQTLLOSNSIVRVQD 86
Db 26 IGCARCECS-----AQKSVCHRRRLAIPEGIPETK----- 61
QY 87 SELCYLANTLDELSONSPSDARCDPHALPOLLSLHEENQRLREHSHRAGLASLOEL 146
Db 62 -----LDLSKRLKLSVNPFEETISYPLLEEIDLSNIIANWPGAFNNLFNLSL 110
QY 147 YLNNQVRIAPRAFSGLSNLRHLNLSNRALDASRWEMLPNLEILMIGNKVDATLD 206
Db 111 RLKGNRLKLVPLGVFTGLSNLTKLDISENKKIYLLDYMFDLNLKSLVEGDDNDLVYISH 170
QY 207 MNFRPLANLRSVLAMNLREISDALEGLQSLSESYDNQALRVPRALQVPGIKF 266
Db 171 RAFSGLSLEQLTEKCNLTAVPTEALSHLSLSLKLKLNINMWYAFKRLFHLKL 230
QY 267 DLNNKPLQVPGDPFANMLHKLGLNNMELVSDKFALVNPBELTKLDITNPNP----- 321
Db 221 EIDWPLDMWPANSVLGLNLTSLVTN-TNLSTVPFLAKHLYLLHNLSTNPISITIE 289
QY 322 -----RLSFHPRAFHLPMQETLMLNNALSLALHOQVESLPLQOE 363
Db 290 AGMFSDLRLQELHIVGAQLTIEPHSFQGLRFLRVLVNSONLLETLEEVNFPSPRALEV 349
QY 364 VGLGNPDRCDVIRMANATGRVRFTEPOSTLCAEPDQRLPVRVPREMTDHLPL 423
Db 350 LSINNPLACRCRLMILIQROPILQF-GGQOPMAGD-----TIRRSKDHSTALSF 403
QY 424 I-----SPRSPPSLQ---VASGESMVLHCLALPEPEEYIWPVPAGLRLPAHAGRRYR 476
Db 404 YFTCKPKPIREKKLQHLVLDGQTVOLECSADGDPQVLSWVPRRRRITTKSNGRA-TV 462
QY 477 YPEGTELEIRYTAEEAGLYTCVAQNILVAGDPKTYSVYV 514
Db 463 LGDGLTEIRFAODDQSGMYVCIASNAAGNDFTTASLTV 500

```

RESULT 14

P70193 PRELIMINARY: PRT; 1091 AA.

ID P70193

AC P70193;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE MEMBRANE GLYCOPROTEIN.
 GN IMG.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10990;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96394313; PubMed=8798419;
 RA Suzuki Y., Sato N., Toyama M., Wanaka A., Takagi T.;
 RT "cDNA cloning of a novel membrane glycoprotein that is expressed
 specifically in glial cells in the mouse brain LIG-1: A protein with
 leucine-rich repeats and immunoglobulin-like domains.";
 RL J. Biol. Chem. 271:22522-22527(1996).
 DR EMBL: D78572; BAA11416.1; -.
 DR HSSP: P56276; ITLK.
 DR MGD: MGI:107935; Img.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00047; Ig_3.
 DR Pfam: PF00560; LRRCT; 14.
 DR Pfam: PF01463; LRRCT; 1.
 DR Pfam: PF01462; LRRNT; 1.
 DR SMART: SM00408; IgC2; 3.
 DR SMART: SM00370; LRR; 6.
 DR SMART: SM00082; LRRCT; 1.
 DR SMART: SM00013; LRRNT; 1.
 DR SMART: SM00369; LRR_Typ; 4.
 KW Immunoglobulin domain.
 SQ SEQUENCE 1091 AA; 119283 MW; A13D0866CE4C203D CRC64;

Query Match 11.8%; Score 442; DB 11; Length 1091;
 Best Local Similarity 23.2%; Pred. No. 6,9e-25;
 Matches 169; Conservative 106; Mismatches 253; Indels 198; Gaps 22;

QY 5 VADLLAWV-----AGATATVVPVWVPCPPQCAC-----QI 37
 DB 14 LARLLMLLLLLQWPESAGAARP-----RAPCAACTCAGNSLDCSGRLATLPRDL 68
 QY 38 RPYVYTRS---SYREATYVD-----CNDLFLTVAPPALPA--GTOTLLLOS 78
 DB 69 PSW--TRSLNLSTYRLSEIDSAFEDLTNIQEVYLSNNELTAPISLGTASIGVSVSLDQ 126
 QY 79 NSIVRDOSELGYLANLTFLDSONSFSDARCDFFALPOLLSEHLNEENLRLDHSFA 138
 DB 127 NKLISVDSQGLKSYSLLEVLDLSSNNITIRSSCPNGRLRIEMLNASKRISLILSGAFD 186
 QY 139 GLA-STQELYLNNHQLYRIAPRAFGLSNLRLHNSNLRAIDSRWEPMLNLEIIMG 197
 DB 187 GLSRSLTLRLSKNRITQLPVKAfk-LPRLTQLDLNRNRIRLIEGLTGGDLSTLEVLRQ 245
 QY 198 GKNVDAIILDMNRPRLANLSTYLACNNLEIDYALEGLQSLESISFYDNOCLARVPR--- 254
 DB 246 RNNISRLTGALFWGLSKMVLHLEYNLSLVEVNSGLYGTALHQLHLSNNSISRIORGW 305
 QY 255 -----RALBOVPGKFLDKLNKNPLQGVGDPANMHLKELGN 293
 DB 306 STCQKHELILSFNNLTRIDESLAEISLTLRLSHNAISIAAGAKGLKSLLEVLD 365
 QY 294 NNEELVSD--KFAVLNPELTAKLDITNNPRLSFIHPRAFHILPOMETLMLNNALSALH 351
 DB 366 HNEISGTLIDTSGAFTGLDNLKLTLPFGN-KIKSVAKRAFSQLESLDEHLNLTGENAIRSVQ 424

QY 352 QQVTESLPNLQEVGLHGNPIRDCVIRW-----ANNT-----GTRVRFI 390
 DB 425 FDAFAKMNKLKELYISSFLDCQOLKWLPPMLGRMLQAFVATCAHPESLKGOSIFSV 484
 QY 391 EPQSLCAEP-----DL 403
 DB 485 LPDSFVCDFFRPQIITQPEPTMAVVGKDIRFTCSAASSSSPMTFAMKDNELANADM 544
 QY 404 QR-----LPREVP-----REKTDCLPLIS 425
 DB 545 ENFAHVRQDGEVMEYTTILRLRHVTGHEGRYOCIIITNHEGYSYSHKARLTAVLP--S 602
 QY 426 PRSEPPSLQVAGSGSMVLHCRALAEPEPELYVTTPAGILRLTPAHAGRRYRYP-EGTLEL 484
 DB 603 FTKIPHDIAIRTGTLARLECAATGHPNPQIAMQKDGTFD-PAARERRMHVMPDDVFEI 661
 QY 485 RRYTKEEGLTCAVQNLVVG--ADTKYSVYVGALLQPGDVE----GQGLRLRVQDTPH 538
 DB 662 TDVKRIDMGVYSCTPAONSAGSVSANATLTLEPISLAVPLEDRVVTGVTVAFOCKATGS 721
 QY 539 YHILSWV 546
 DB 722 PTPRTTWL 729
 RESULT 15
 ID 09HCJ2 PRELIMINARY; PRT; 640 AA.
 AC 09HCJ2;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE KIA1580 PROTEIN (FRAGMENT).
 GN KIA1580.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20450683; PubMed=10997877;
 RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.
 XVIII. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro.";
 RL DNA Res. 7:273-281(2000).
 DR EMBL: AB046800; BAB13406.1; -.
 DR HSSP: P22888; ITUM.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000483; LRR_Cterm.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00047; Ig_1.
 DR Pfam: PF00560; LRR; 9.
 DR Pfam: PF01463; LRRCT; 1.
 DR Pfam: PF01462; LRRNT; 1.
 DR PRINTS; PR00019; LEURICHKPT.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00408; IGc2; 1.
 DR SMART; SM00410; IG_Like; 1.
 DR SMART; SM00370; LRR; 6.
 DR SMART; SM00082; LRRCT; 1.
 DR SMART; SM00013; LRRNT; 1.
 DR SMART; SM00369; LRR_Typ; 8.
 KW Immunoglobulin domain.
 FT NON TER 1
 SQ SEQUENCE 640 AA; 71949 MW; 6058974872636838 CRC64;

Query Match 11.7%; Score 438.5; DB 4; Length 640;
Best Local Similarity 25.5%; Pred. No. 5.8e-25;
Matches 148; Conservative 84; Mismatches 231; Indels 117; Gaps 12;

```
QY 2 RLVAFLLLAVAGATATVPVPHVHPQACQIRPWYTPRSSYREATVDCNDLFLT 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 20 RALFDPFLVVLALQLLVAGLVRAQCPSVCSCS-----NQFSKVICVRKNLR 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 AVPPALPAGTQTLLQNSIVRDQSEGLATLUTELDLSQNSFSDARDCDFHALPQLLS 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 69 EVPQISTNTNLTLLHENOIQIIVKNSFKLRHLEITQLSRNHI RTEIGAEVGLANLNT 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 122 LHLBNQTLRLDHSFAGLASLOELYLNHQLYRIAPRAFSGLSNLLRLHNSNLLRAID 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 129 LELFDNRLLTTPNGAFVYLSKRLKELMLRNPNIESIPSAFNRIPLRLDLGE----- 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 SRWEEMLPNLEILMGKNKVDAILDMNFRPLANRSLVLAGMNLREISDYALEGLQSLSES 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 -----LKRLSYISEGA-----FEGLSNLRYLNLAMCNLREIPN----- 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 LSFYDNOIARVPRRALDQVPGKFLDLKNPQVRGPGDFANMLHLKELGNNMELVSI 301
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 215 -----LPLIKDELDELDSGNHLSAIRGSPQGLMHLOKLMIGSQ----- 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 302 DKFALVNPETLTKDITNNPRLSFIHPRAFHHLPMETLMLNNALSAHQTVESLPNL 361
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 255 -----IOVIERNAFDNLOSLVEINLAHNNLTLLPHDLFTPLHL 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 362 QEVGLHGNPIRCDCVIRMANATGTRVRIEPOSTICA---EPPDLQRLPREVPREMT 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 ERIHHNHPWNCNDIWLSS--WIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFT 350
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 418 DHCLPLISPRSPSLQVASESMVLCRALAEPEPEIYVTPAGLRLTPAHNGRXYRY 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 351 CYAPYIVP---PADLVNTEGMAELKCRA-STLSVSWITPNGVTMTHGAYKVRIVAVL 406
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 478 PEGTLELRRTAEEAGLYTCVAQNLVGADTKTVSVVVGRA-----LLOPGR 523
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 407 SDGTINFTNVTYQDTGMTCVNSYGNJTASATLANTTAATTPESYFSTVETMEPSQ 466
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 524 DEGGGLELRVOETHRYHLLSWTPPNTVSTNLTWSASS 563
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 467 DEARTDNNVGPTP---VVDMEF--TNVTTISLTPOSTRS 500
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: August 31, 2002, 14:40:33
Job time: 260 sec

